# Journal
# of
# Applied Physics

---

VOLUME 18

JANUARY DECEMBER, 1947

---

*Published by the*
AMERICAN INSTITUTE OF PHYSICS
Incorporated

# INFORMATION FOR CONTRIBUTORS

## Manuscripts

Send manuscripts to Elmer Hutchisson, Case School of Applied Science, Cleveland 6, Ohio.

Papers must be in English, **typewritten double spaced** with wide margins on a durable white paper, preferably lettersize. **The original copy,** not the carbon, should be submitted.

For general style, spellings, abbreviations, and form, recent issues of this journal should be consulted.

## Abstracts

**An abstract must accompany each article.** It should be adequate as an *index* and as a *summary.* As an index it should give all subjects, major and minor, concerning which new information is presented. As a summary it should give the conclusions of the article and all numerical results of general interest.

## Mathematical Expressions

Make all expressions clear to the typesetter. Identify in the margin of the manuscript Greek letters and unusual symbols.

Use fractional exponents instead of root signs.

**Avoid complicated exponents and subscripts.** Should it be necessary to repeat a complicated expression a number of times it should be represented by some convenient symbol.

The solidus (/) should be used wherever possible for fractions.

## References

References should appear as footnotes, numbered consecutively, and arranged thus:

[1] A. B. Smith, Phys. Rev. 41, 852 (1932).

[2] G. R. Harrison, *Atoms in Action* (William Morrow & Company, New York, 1939), first edition, p. 30.

## Figures and Tables

**A separate sheet listing captions for all figures must accompany manuscript.** Each table should have a caption. These in all cases should be complete in themselves so as to make the data intelligible to the reader without reference to the text. Each figure and table should be cited in sequence in the text.

## Line Drawings and Photographs

Plan figures for approximately one-half reduction. Wherever possible arrange figures so that after reduction they will not be wider than **three inches. Lettering must be large enough to be readable after reduction.**

**Line drawings must be made with black India ink on white paper or tracing cloth. Coordinate paper is not desirable, but if used must be blue-lined and all coordinates to be reproduced drawn with India ink.**

Drawings which cannot be used will be returned to the author.

Photographs should show clear black and white contrasts.

Avoid round and oval photographs.

## Alterations

**Take great care that each manuscript is typographically accurate.** A limited number of alterations in proof is unavoidable, but the cost of making extensive alterations after the article has been set in type will be charged to the author.

Proof and all correspondence concerning papers in the process of publication should be addressed to the Publications Manager, American Institute of Physics, 57 East 55 Street, New York 22, New York.

# INFORMATION FOR SUBSCRIBERS

The JOURNAL OF APPLIED PHYSICS is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. It aims to be of service not only in physical laboratories, but also in the laboratories of industrial, chemical, geological, geophysical, meteorological, radio, and similar concerns. The JOURNAL OF APPLIED PHYSICS publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics.

**Subscriptions** and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

### SUBSCRIPTION PRICE

| | U. S. AND CANADA | ELSEWHERE |
|---|---|---|
| To members and associate members of the American Institute of Physics | $5.00 | $5.70 |
| To all others | 7.00 | 7.70 |

### BACK NUMBER PRICE

Complete set: Vol. 1, July, 1931 –Vol. 17, 1946–$108.50
Yearly back number rate: $7.70.
Single copies: $0.70 each.

**Advertising rates** will be supplied on request.

**American Institute of Physics**

**57 East 55 Street** **New York 22, New York**

# CONTENTS

## January, 1947

## February, 1947

MARCH, 1947

APRIL, 1947

**Letters to the Editor:**

MAY, 1947

*Special Issue on High Polymers*

June, 1947

July, 1947

**Letters to the Editor:**

AUGUST, 1947

**Letters to the Editor:**

SEPTEMBER, 1947

**Letters to the Editor:**

## October, 1947

Letters to the Editor:

## November, 1947

Letters to the Editor:

## DECEMBER, 1947

**Letters to the Editor:**

Journal
of
Applied Physics

Volume 18, Number 1 January, 1947

# European Electron Induction Accelerators

BY HERMAN F. KAISER
*U. S. Naval Research Laboratory, Washington, D. C.*

Betatrons, induction electron accelerators for the production of high energy $X$-radiation, were under development in Germany during the war and to a limited extent at present. This work is reviewed for the most part and details are given on the constructional features of 6- to 15-million volt betatrons and on the theory and design of 15- and 200-million volt betatrons. The smaller units, especially the Siemen's 6 Mv, are quite successful and more advanced than comparable American units. While no large machines were actually built, the projected 200-Mv Wideroe design introduces new features of value. Betatron patent ideas are reviewed.

## I. HISTORY OF BETATRON DEVELOPMENT ON THE CONTINENT

THE first published work on electron induction acceleration to appear in Europe was the doctoral dissertation of Rolf Wideroe[1] of 1927 which was published as a paper[2] the following year. It is known, however, that a patent application for an electron induction accelerator was filed in Germany by J. Slepian (Westinghouse) soon after the filing of an application[3] in the United States in April, 1922. Because of a misunderstanding of the principles involved by a Dr. Smidt of the German Patent office who persisted in quoting the text of Abraham and Becker in support of his contention that it was impossible to both accelerate and guide an electron by a magnetic field at one and the same time, it was not granted until 1928. Whether or not Wideroe had knowledge of this patent is not known. At the same time experiments were being carried out by Breit and Tuve[4] at the Carnegie Institute in Washington which it is now known were correct in principle and might have been successful if pursued further. In 1929 a work by E. T. S. Walton,[5] following a suggestion by Lord, Rutherford, appeared in England. In this he discovered independently the flux conditions announced by Wideroe and made a further contribution on the theory of the magnetic fields required for assurance of a stable orbit. Successful experimental results unfortunately were not obtained.

Developments on the continent have been based on Wideroe's work and on the Steenbeck patents of 1937[6] in which the necessary conditions of Wideroe are set forth and supplemented

---

[1] Dissertation, Tech. Hochschule Aachen, 29 Oct. 1927.

[2] Archiv. f. Elektrotechnik 21, 387 (1928).

[3] U. S. App. Ser. 548, 630, April 1, 1922, U. S. Pat. 1,645,304, Oct. 11, 1927.

[4] Carnegie Institution Yearbook **27** (1927–1928), p. 209.

[5] Proc. Camb. Phil. Soc. [9] **25**, 469 (Oct. 1929).

[6] DRP 656,378 and 698,867; U.S. Pat. 2,103,303, Dec. 8, 1937.

by the condition for producing axial stability by curvature of the guide field flux—a condition which was discovered independently in the United States by Breit and Tuve and later by Lawrence and Livingston for the cyclotron. An induction accelerator was built and tried by Steenbeck, who recorded production of 1.8-Mev radiation. Following Kerst's[7] publication of successful production of 2.5-Mev radiation at intensities of 15–100 milligrams of radiation, Steenbeck published a belated note on his work in 1943,[8] describing results obtained in 1935 and 1936.

The work of Kerst was a great incentive to the continental workers interested in this field, particularly Steenbeck and Wideroe. During the war the latter was the more active in this work, but Steenbeck set into action further work on this subject by Siemens Reiniger. Wideroe and the group that came to be associated with him in the wartime German betatron work were not in sympathy with the Nazi cause but were induced to continue such work for scientific and other considerations.

An unpublished investigation[9] interrupted by the war was carried out by James L. Tuck working in collaboration with L. Szilard at the Clarendon Laboratory in Oxford. It is now known that this investigation came very near to success.

The betatron development was carried out by three separate groups. One was the Megavolt Versuchsanstalt (MVA) or Megavolt Research Association which worked under the guidance of Dr. Wideroe. Siemens and the A.E.G. also had betatron projects. MVA laid plans for 15-, 30-, and a 200-Mv accelerators. The first was begun early in 1944 and brought to successful completion in the fall of the same year. The second was merely tentative. The design of the 200 Mv was prompted by news that the General Electric Company (U.S.A.) was constructing a 100-Mv betatron. The design of the 200-Mv accelerator was completed about this time and the construction work was turned over to the Brown-Boveri firm in Heidelberg. Occupation in Germany, however, came before any construction work could be begun.

The basic ideas underlying the design of the 15- and 200-Mv induction accelerators were published by Wideroe in two papers.[10] These discuss the special problems involved such as the axial and radial stabilization by properly designed guiding fields, injection and extraction of electrons, the useful load of the apparatus, the maximum electron filling of the acceleration chamber, dependence on injection voltage, use of auxiliary lens systems especially for high radiation energies, the effects of gas molecules, technical transformer construction for high energy production and finally the fields of application of the betatron; cancer treatment, radiography, and nuclear physics.

In collaboration with the design work of Wideroe, a considerable amount of theoretical work was carried out by Touschek which was known to have been of invaluable aid in the development of the 15-Mv accelerator. Further theoretical work has been done by Touschek on the starting of electrons in the accelerator. Some of this work is along the lines initiated by Kerst and Serber[11] whose work was known to Touschek. Recent work of his, it is believed, may give an explanation of the discrepancy between the Kerst-Serber theory and experimental results.

The wartime betatron work of the Siemens Reiniger firm was based on the work of Steenbeck whose 1940 patent is mentioned above. An American license for its use by the General Electric Company was requested from Siemens shortly before the outbreak of war in 1941. The results published by Kerst[11] renewed an interest which had lagged and in 1942 (December) designs were under way for construction of an electron induction accelerator. Dr. Steenbeck gave the basic specifications which were incorporated into a design at Siemens and discussed in a conference on December 15, 1942 held together with Dr. Ganswindt, Dr. Kurt Bischoff, Dr. J. Patzeld, and Dr. Konrad Gund all of Siemens. At this time was also discussed a design of Dr. Gund in which he proposed an accelerator with a fixed magnetic guide field, surrounding an alternating magnetic field passing through a central core, a

---

[7] Phys. Rev. 60, 47 (1941).

[8] Naturwiss. 19/20, 234 (1943).

[9] Private communication from D. W. Kerst.

[10] Wideroe, Part I, Archiv. f. Elektrotechnik 37, 391–408 (1943). Part II remains in proof form.

[11] D. Kerst, Phys. Rev. 58, 841 (1940); Phys. Rev. 59, 110 (1941); and 60, 47–53 (1941).

scheme originating from a paper of W. W. Jasinsky[12] in which the field-radius conditions are erroneously set forth. These are corrected in the design mentioned and several other ideas set forth. A patent application was filed by Siemens in July 1942[13] covering these and also premagnetization or bias field operation. Ideas for operation with biased fields were also evolved by Kerst at about this time.[14]

From the design worked out in the above conference and later meetings a successful 6-Mv betatron operating on 550 cycles was developed. A more detailed description of this unit will be given later.

## II. ACTIVITIES OF THE MEGAVOLT RESEARCH ASSOCIATION (MVA)

### A. Description of the 15-Mv Induction Electron Accelerator

*General Considerations*

Theoretical curves of the absorption coefficient as a function of electron energy have a minimum which lie between 3–10 Mev for most metals. A peak energy of 15 Mev was, therefore, chosen to bring the effective energy in the neighborhood of 7–8 Mev. (It has since been shown that 20 Mev satisfies these requirements better.) A theoretical sixfold increase in the range of radiographic thicknesses was expected. It was also desired to use the betatron for medical purposes but not too much was expected except for an improvement in the depth dose condition due to a lowered absorption coefficient. The desirability was recognized of using the betatron for the production of high energy electron beams in which the dissipation of energy by absorption in human tissue may be localized more effectively. This problem was under consideration but no practical results had been realized.

*Constructional Details, Core*

The radius of the electron equilibrium orbit was chosen as $r=14$ cm. The diameter of the induction pole was chosen as 220 mm (11-cm radius) and the air gap in the poles was made 2 cm. The average separation of the guide poles was chosen as 6 cm and the contour of the guide pole faces was determined from magnetic measurements and from measurements made by means of the electrolytic trough. This process was continued until the guide field had the properties required. The guide poles were so arranged with respect to the remainder of the apparatus that, as a result of the saturation of the guide field, the latter does not increase in proportion to the accelerating field from a certain point in the cycle onwards. This type of construction has some obvious advantages. The space between the acceleration pole and guide poles allows the use of flux biasing windings. It also has the advantage that no orbit expansion equipment is required, the electron striking the target upon saturation of the guide field and this without the attendant loss of energy which is present when the electron has to spiral into a smaller radius to strike a target. It may be noted, however, that this feature may still be obtained in Kersts' form of construction in which acceleration and guide fields arise from the same pole piece by a special reduction of the amount of iron at the edge of the pole piece. The approximate data for the main components (which are slightly above the initial estimates of Dr. Wideroe) are:

| | | |
|---|---|---|
| Iron weight | 1000 kg | 2200 lb. |
| Copper weight | 200 kg | 440 lb. |
| Total losses | 2.5 kw at 50 cycles | |
| Dead load | 250 kva | |
| Driving voltage | 6 kv | |

The details of core and pole assembly are shown in Fig. 1. The yoke is constructed from laminated iron of thickness equivalent to 29-gauge, 0.014-inch stock of approximately ½ watt per pound core loss (10,000 lines and 60 cycles). The lamination stack is not divided, i.e., no provision is made for cooling the core by means of air ducts. It was found that the heating was not excessive. The vertical members of the yoke, in contrast to American design, extend the full height of the apparatus, and the horizontal members are held between these by a special arrangement of clamping frames. This produces four vertical "shim" gaps. On the upper side these have means for adjustment of the height and level of the upper transverse member (yoke). This part carries the accelerating pole by means of six bolts circularly arranged. By means of a supporting ring, it also carries the guide pole and

---

[12] Archiv. f. Elektrotechnik **30**, 500 (1936).
[13] Siemens Akt. 151,465, VIII c/211g.
[14] D. W. Kerst, Phys. Rev. **68**, 233 (1945).

FIG. 1. View of 15-Mev betatron at Wrist.

coil boxes. The yoke laminations are clamped by transverse bolts through insulating tubes.

### *Acceleration Chamber*

Glass acceleration chambers were made by the glass-blowing laboratory of the C. H. F. Müller firm. These have an approximately elliptical cross section. The inside walls are coated with graphite or silver coatings and the tubes have either two or three radial openings with stems on the periphery for evacuation or for mounting the electron injection assembly. Vacuum was produced by an oil diffusion pump backed by a three-stage mercury diffusion pump with a liquid air trap separating the two. The mercury diffusion pump was backed in turn by a rotary oil pump. This unnecessary complication of pumps was caused by the scarcity of available and usable equipment.

The vacuum was measured by means of a Phillips vacuum meter placed between the two diffusion pumps and by means of an ionization gauge sealed to the acceleration chamber on the injection stem. It was noted that operation was not possible at pressures above $10^{-5}$ mm Hg and steady yields were obtained only at $10^{-6}$-mm pressure.

The electron gun or injector is shown in Fig. 2 and follows in principle the design introduced by Kerst. It consists of a tungsten filament (oxide coated electron sources were also used) surrounded by a semi-cylindrical Wehnelt or Faraday cage and of two anode plates insulated from each other. The cage serves, on application of a potential between itself and the filament, to concentrate the electron stream. The split anode, on the two halves of which different potentials could be applied has, ignoring for a moment the necessary acceleration of electrons, a purpose similar to that of the deflection plates of a cathode-ray tube. It is thus possible by applying various potentials to these plates to give the electrons various initial projection angles in the plane of the orbit. The necessary correction of the electron beam in the vertical direction is obtained by a small rotation of the whole injector system or stem on its ground glass joint. A small bar of tungsten is fastened to the innermost anode plate with its long axis vertical to serve as a target for x-ray production.

The filament is heated continuously. The injection voltage is applied only for a short time at the beginning of the acceleration period by means of a special arrangement to be described below, partly to charge the entire injector structure to as high a voltage as possible and partly to avoid filling the electron accelerator chamber with stray electrons.

### *Connections*

Figure 3 shows the entire diagram of connections. Alternating current of 6 kv is generated by a regulating transformer and supplied to the driving circuit in which the dead load of the circuit is compensated by a capacitor bank, the

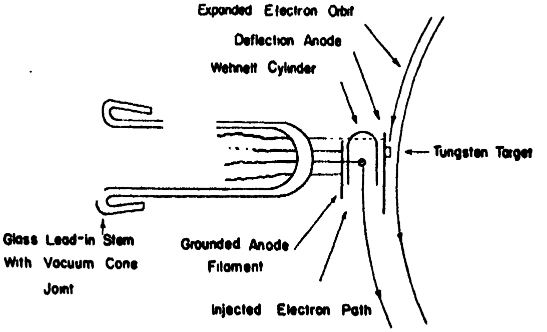


FIG. 2. Electron injector gun for 15-Mev betatron (schematic).

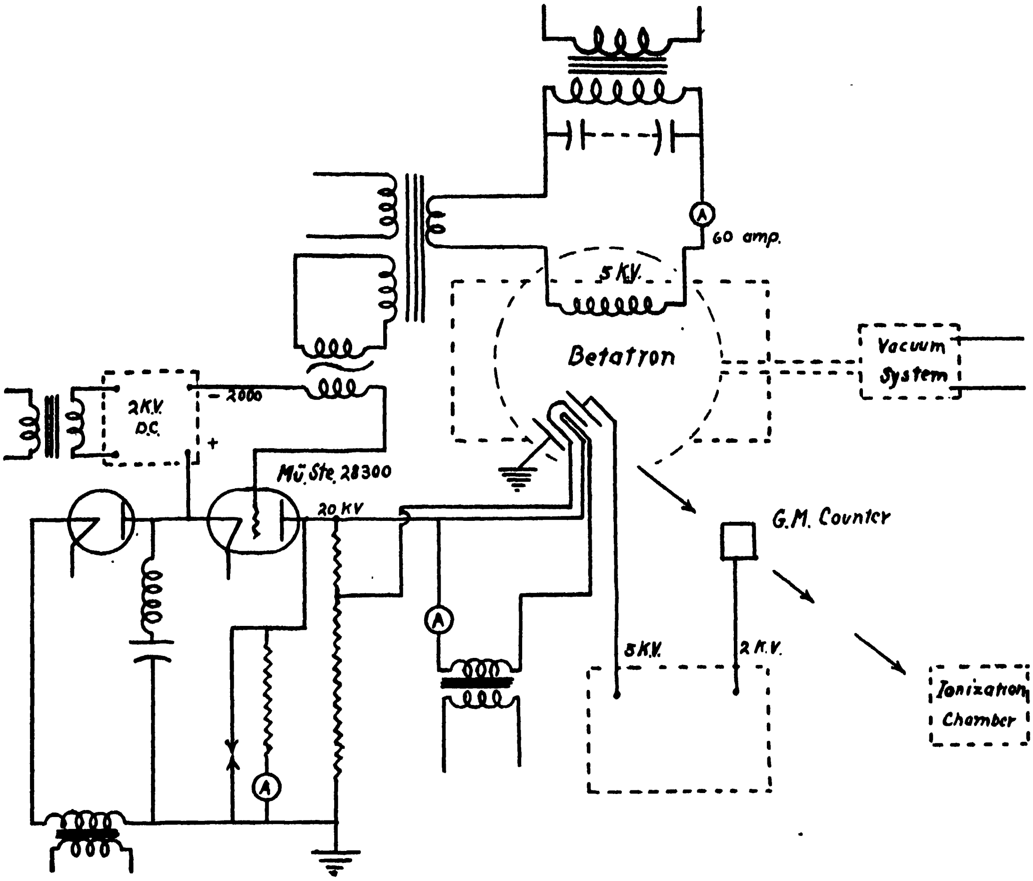


FIG. 3. Diagram of connections for 15-Mev betatron.

capacity of which may be varied between limits. The injection voltage is produced by H.V. equipment from which the potential may be applied at the proper instant near the beginning of the acceleration period by means of a special grid controlled high voltage rectifier tube. The driving circuit of the betatron passes through a small utility transformer which in turn operates a peaking transformer for the production of a sharp positive pulse to remove momentarily the large negative potential on the grid of the control valve. In earlier experiments, a mechanical contactor driven by a synchronous motor was used to apply the tripping pulse but this gave much trouble due to fouling of contacts. The circuit diagram also shows a 5-kv constant potential circuit used for applying the deflection voltages to the split anode, the innermost plate of which is connected together with the acceleration chamber coating to ground.

*Operation*

In the early stages of operation radiation is detected by means of a counter tube which of course becomes useless when the machine is brought into adjustment and radiation of considerable intensity is produced. Ionization chambers are then used and depended upon for measurement of the radiation intensity. With shielding the counter tube may still be used for indicating the point in the cycle at which radiation is emitted directly upon an oscilloscope in which the horizontal or time base is supplied by the a.c. driving voltage.

In operation the vacuum is checked, the driving circuit is energized, and the phase of the injection pulse is determined and corrected by means of the phase shifting device. On starting the counter and cathode-ray tube, a radiation output is generally observed. By rotation of the ground glass joint of the injector, by application of a suitable deflector voltage, by changing the position of the acceleration chamber in the machine, and also by adjustment of the air gap in the guide field it is possible to gradually bring about an optimum yield. Then by keeping these variables constant quite reproducible results could be depended upon which were then only influenced by the vacuum conditions.

To check the quality of the radiation produced, various methods were resorted to: (a) observation of the output radiation pulse on the oscilloscope allowed an estimate of the time of acceleration of the electrons and a deduction of their maximum energy which in general was around 13–15 Mev, (b) absorption measurements on the radiation which, however, are somewhat uncertain because of the shallow minimum in the absorption *versus* energy curve, (c) rapid dying out of radiation intensity on reducing the amplitude of the exciting current, and (d) production of artificial radioactivity by irradiation of various materials. These phenomena are especially pronounced in the case of copper and silver.

After the expected quality of radiation had been established, attention was directed to improving the intensity of the radiation. It was hoped to attain a maximum electron filling of the acceleration chamber such as could be expected from the following formula deduced by Wideroe for the effect of space charge*:

$$i=\frac{4\pi n^3(\Delta r)^3(1-n)f^2rB_{\max}}{c^2(-dr)}\text{ amp.}$$

$\Delta r$: radial oscillation amplitude,
$dr$: change in $\Delta r$,

in which $r$ is the radius of the orbit, $n$ the exponent in field-radio relation, $f$ the frequency, $B_{\max}$ the maximum guide field strength, and the other quantities their usual significance. This formula indicates a value which is usually about ten times that observed experimentally. In operation yields ranging from 30 grams to 1 kilogram radium equivalent were obtained. It was believed that this discrepancy is caused by the release of gases by bombardment of the walls of the acceleration chamber and that better results could be obtained only by use of baked out and sealed-off acceleration chambers.

### *Program for Further Development*

The program of the Megavolt Versuchsanstalt as described in June 1945 was: (a) further work to increase the radiation yield of the 15-Mev betatron and (b) to build as soon as possible a duplicate of this equipment for use in medical research to allow the first unit to be used solely for technical development. This was, of course, the wartime program. At present scientific research is strictly limited by the military government of the occupational forces.

Under (a) it was planned to make further experiments on acceleration chambers fitted with electron lens systems which might, in view of the greater stabilization forces made available by their use, bring about an increase in radiation output intensity. It was also planned to study the tangential electron injection technique of Wideroe once more because of the possible advantage accruing from the absence of an injector structure in the acceleration chamber; also to study the problem of releasing a beam of electrons from the apparatus, a subject of considerable importance.

Another intention was the design and construction of a 30-Mev unit for use in nuclear research. This project was to have been undertaken at an early date and some of the design calculations had been made along with those for the 15-Mev betatron.

It was felt that an apparatus of equal size but operating on the premagnetization scheme of Wideroe would be useful in testing such questions as to whether or not the final energy of the accelerator would be affected by the radiation damping as predicted by Iwanenko and Pomerantschuk.[15]

A final project was the design and construction of a 200-Mev accelerator using the premagnetization scheme. This will be described below.

## B. Description of the 200-Mev Induction Electron Accelerator

### *Preliminary Considerations*

The design of the 200-Mev betatron was the outgrowth of some preliminary design work carried out by Wideroe for a 100-Mev unit.[10] This was based on his experience with the con-

---

* A brief derivation is given in Appendix I.

[15] Phys. Rev. 65, 43 (1944).

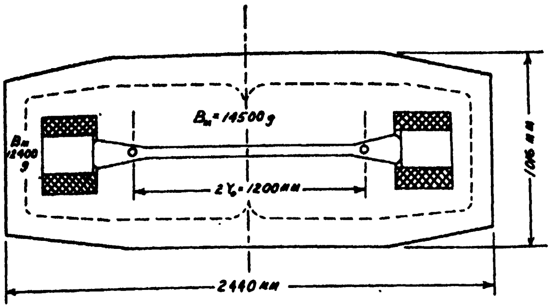


FIG. 4. Design for a 100-Mev betatron.

struction of large transformers by the Brown-Boveri firm with which he was associated. An economical design for transformer construction used by this firm consists of a radial construction in which the radial laminations are welded at the top and bottom. In this type of construction an exactly circularly shaped iron section is obtained and the required guide pole shapes are easily designed and adapted. The necessary cooling ducts are likewise easily incorporated into the core. The minimum amount of iron is used if the yoke is continuous as in a shell transformer. This lends itself to a rugged construction having a small yoke height and large cooling surface. It also offers a certain amount of radiation protection by shielding the acceleration chamber, access to which is obtained by apertures cut into the yoke.

A preliminary design for a 100-Mev unit is shown in Fig. 4. Here yoke, core, and iron return form a common iron unit constructed from radial laminations. By removal of the ten-ton upper half, access is had to the windings and the acceleration chamber. The core is crossed with air ducts to remove the 13.6-kw iron losses (0.35-mm laminations with a loss of 1.1 w/kg are assumed). The maximum induction of the core is 14,500 gauss for a stacking factor of 77 percent. The yoke and iron returns are able to dissipate 28 kw by surface cooling. In these parts the stacking factor is taken as 90 percent and the maximum inductions chosen are 14,500 and 12,500 gauss, respectively. The total weight is 20 tons.

The air gap for an equilibrium orbit of radius =60 cm is 5 cm and about 20,000 effective ampere turns are required to produce in it a guide field of 5600 gauss. To accomodate the exciting winding (46.5 kg of copper, assuming a current density of 2 amperes/mm$^2$), a window of 800 cm$^2$ is provided. Thus a space factor 12.4 percent is obtained, allowing insulation of the winding to 30–50 kv. This voltage is so chosen that it will be the most economical one considering cost of capacitors. The copper losses of the transformer will be 5–6 kw and the required dead load 7500 kva. The purely constructional costs aside from development and research are estimated at 80,000 RM for the betatron, 120,000 RM for the capacitors, thus making a total cost of 200,000 RM. Using this 100-Mev design as a basis the weights, costs, dimensions, etc. of still larger machines may be estimated as has been done by Wideroe. Of interest here is the estimation for a 200-Mev machine which would have the following characteristics: core diameter 4.89 meters, core height 2.04 meters, weight 160 tons, core losses 400 kw, radiation load 400 kw, and cost about 1,120,000 RM. It is apparent that unless further refinement were introduced betatrons of still higher energy would be extremely costly.

### *Improvement of Yield by Means of Auxillary Lenses*

It was the hope of Wideroe and his group to increase the electron filling of the acceleration chamber by means of electro-static or electro-magnetic lenses. In the 100-Mev unit it was planned to fill the air gap as completely as possible with the acceleration chamber and to provide multiple injectors and guide fields on its periphery. Allowing four guide fields for each oscillation of the electron, 28 cathodes will be required for seven oscillations per revolution. With an average electron beam of 1 ma ($V_i$=300 kv) the frequency of the injection voltage will be 17,000 cycles for a 100-Mev betatron operating at 60 cycles. This gives a charging period of $2\times10^{-6}$ second per cycle. The average charging current for these times must therefore be one ampere and requires, assuming an efficiency of 10 percent, a total cathode emission surface of 500 mm$^2$. For 28 cathodes this amounts to 18 mm$^2$ per cathode. Cathode windows will also have to be of this order of size. To prevent any electronic charges from building up on the acceleration chamber the outside must be neutralized and grounded.

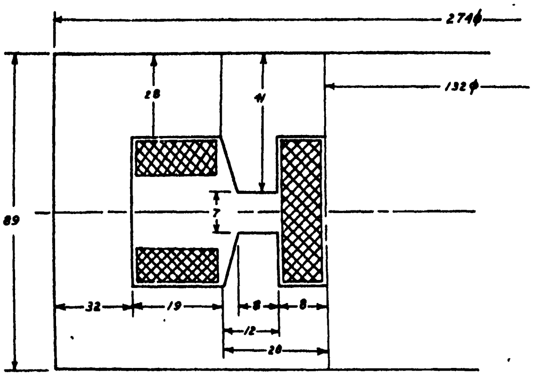


FIG. 5. Schematic construction of 200-Mev betatron (Wideroe).

### *Projected Design and General Description of 200-Mev Betatron**

As mentioned above it was evidently necessary to build a 200-Mev betatron with further refinement of design to keep the size and weight within reason. A first step in this direction was the elimination of the air gap of the 100-Mev design. Even with this feature the diameter of the machine would still be about three meters but a very considerable reduction in the weight was possible. Late in 1944 ideas had been crystallized and the Brown-Boveri firm which is skilled in the construction of large transformers was contracted in August. By mid-October a first rough design had been drawn by the firm which was discussed at length in Heidelberg. In this conference the Brown-Boveri firm was represented by its director Dr. Meyer-Delius, Chief Construction Engineer Otto Weiss, and Dr. Helmut Boecker, detail designer. The Megavolt Versuchsanstalt was represented by Dr. Wideroe and Dr. Kollath. The construction of models and a small experimental unit for the study of special problems was also considered at this time. Exact calculations were to be completed before the end of the year. In November, the MVA received some questions and designs to which reply was made by letter. In January, the first plans for construction were complete and the MVA was informed but it was too late for another conference and discussion of these plans. Since that time, the MVA had no contact with BBC and all work on this project came to an end with the occupation of Germany early in 1944.

The conference held at Heidelberg evolved a design that had the following general characteristics, which may be seen in Figs. 5 and 6. A radially laminated core of 1320 mm diameter was to be surrounded by ring-shaped radially laminated parts of about 200 mm width. These carry the guide poles which have an average separation of 7 cm and also the compensating windings. About these are placed the main exciter windings, divided above and below the median plane. To complete the magnetic induction circuit four mantels are provided 90 degrees apart on the circumference. A difficulty arising from this feature will be discussed below. The preliminary problems of the original design included the purely mechanical construction of the laminated parts of the machine, the fastening of the guide poles to the machine, the division of the transformer to facilitate installation of the acceleration chamber, pumping and electrical connections, and adjustments to the guide poles. It was planned to purchase about 40 tons of 0.35-mm transformer sheet (29 gauge) having a core loss of about 1.1 watts per kilogram.

The upper pole shoes were to be capable of being raised, and the four mantel yokes were to be displaced radially to allow access for leads, etc. A special difficulty was in the evacuation arrangements and in the cooling of the exciter windings. It was proposed to mount the inner winding directly adjacent to the core in a stationary oil case.

### *Final Form Decided Upon*

The general features of the final form evolved in the design work are shown in Fig. 6. In vertical section is shown the central core with its cooling system consisting of a number of parallel helically shaped cooling ducts winding through the body of the core and connecting at the upper and lower ends of the core to the compressed air inflow and return ducts *5*. The core has a stacking factor of about 70 percent. By means of ledges about its periphery, a ring of radially laminated pieces making up the pole shoes is held in proper position, and so that the upper pole shoe ring may be

---

* More detailed design calculations of Dr. Wideroe for both the 15- and 200-Mev units may be found in the author's Naval Technical Mission T.R. 331-45. (Now released to Dept., Commerce Publications Board.)

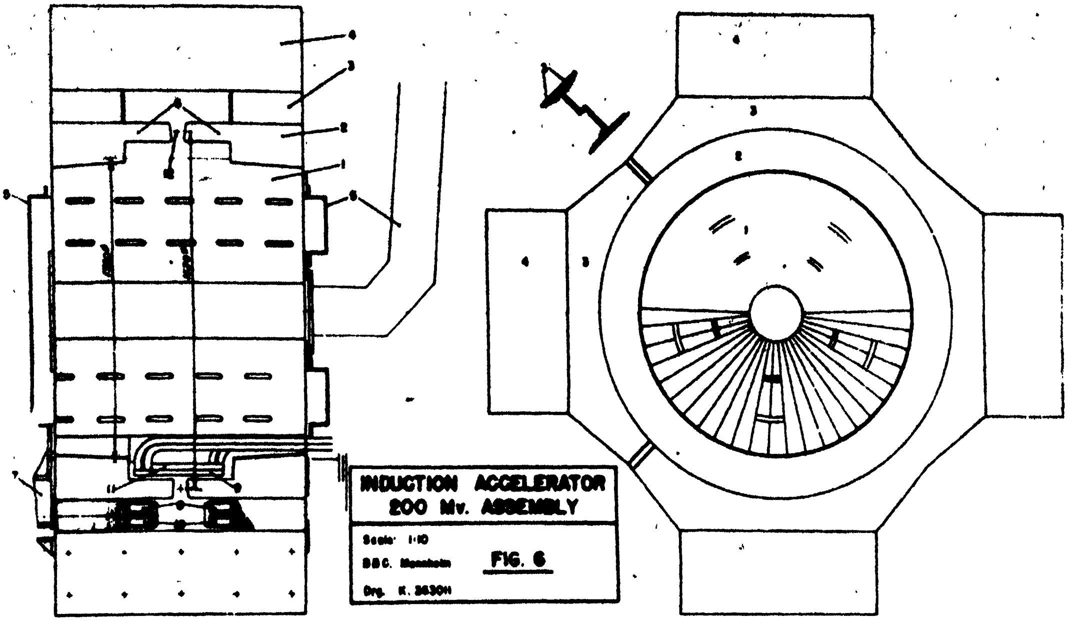


Fig. 6. Assembly of 200-Mev betatron.

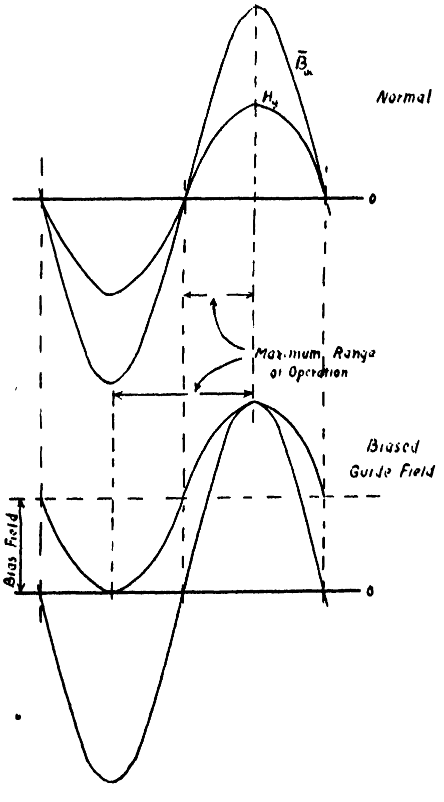


**Fig. 7. Principle of operation with biased guide field.**

raised together with the connecting ring *3*. This carries on its lower side the main a.c. exciter winding and the d.c. premagnetization winding. The original intention to complete the gapless magnetic circuit by means of the four mantels led to the difficulty that the uniformity of the magnetic flux in the guide poles will be disturbed by the locking of the mantels to the guide poles at only a few points. To avoid this difficulty, it was proposed to add an additional ring *3* (horizontally laminated) which serves the purpose of smoothing the flux from the core into the mantels *4*. This is electrically broken between adjacent mantels with a zig-zag joint to avoid an electrically continuous circuit about the flux passing to the guide poles. The figure shows the principal and compensating windings *8* and *9*, and the direct-current premagnetizing windings *10* and *11*.

By elimination of the air gap in this design a large central accelerating flux is produced by the main windings. These windings also serve to produce the guide field which at all times has to bear a fixed relation to the induction flux in order to maintain a stable electron orbit. In this design it is furthermore proposed to premagnetize the guide field so that it will be operative over nearly

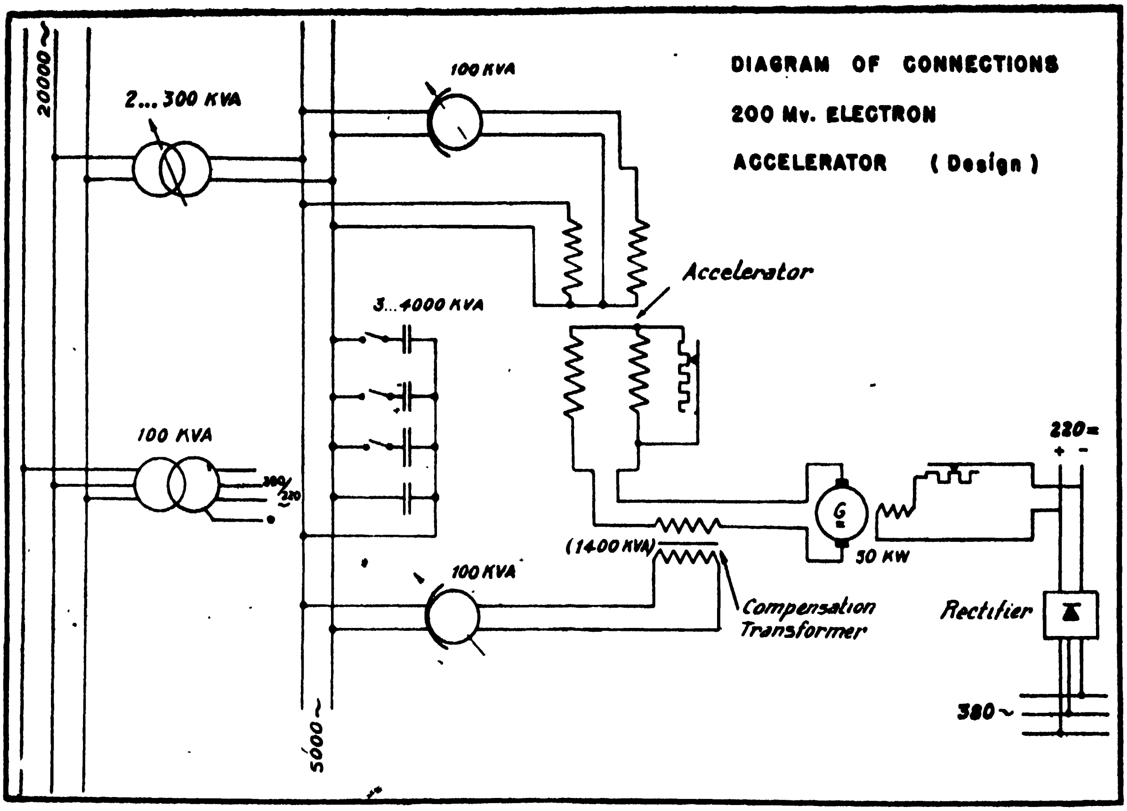


FIG. 8.

a half a cycle instead of a quarter of a cycle, thus permitting the induction field to accelerate the electrons for more than a quarter of a cycle. This principle is illustrated in Fig. 7, and has recently been discussed in the literature.[14,16]

The air gap of the earlier design provided the necessary relationship between the induction and guide fields by insuring their proportionality. Since the use of it leads to unnecessarily large power requirements, an attempt is made to provide the necessary relationship by a winding in the inner coil space (compensating winding) on which is impressed a fixed potential of the corresponding magnitude. For practical purposes the main magnetizing winding and its compensating windings may be connected to the same driving source and the flux *vs.* induction relationship may be regulated by proper proportioning of the windings of the two coils. For fine adjustment a regulation transformer may also be connected into the circuit. Figure 8 shows the manner in which these windings are to be energized from the power sources and the necessary compensating transformer connection to neutralize the a.c. voltage induced in the d.c. windings.

For this design the following windings were planned (see Fig. 6):

1. An alternating current winding placed outside of the guide pole.
2. A compensating winding correspondingly placed about the induction pole to depress the a.c. magnetization of the core to such an extent that the average induction of the core is only twice that of the induction at the center of the guide pole (fulfillment of the necessary 1:2 relation).
3. A premagnetization winding for the guide pole of which the induction is to vary from zero to a maximum.
4. A compensating winding corresponding to the above which, while in principle is not necessary, is used to prevent an undue loading of the machine by a d.c. magnetization with undesirably high induction and saturation characteristics.

Further general electrical characteristics on the design are:

---

[16] W. F. Westendorp, J. App. Phys. 16, 657 (1945).

| | |
|---|---|
| Total induced flux | $105 \times 10^6$ maxwells |
| Induction in core | 15,000 gauss |
| Induction in yoke | 13,900 gauss |
| Induction in mantels (4) | 12,200 gauss |
| Induction in guide pole | 15,000 gauss |
| Total iron weight | 26 tons |
| Volts/turn at equilibrium orbit | 366 volts |
| Maximum electron energy | 200.5 Mev |
| Operating power, driving circuit | 7900 kva |
| Operating power, compensating windings | 4800 kva |
| Resulting net load | 3100 kva |
| Iron losses about | 67 kw. |

Details on the exact form of the guide poles were not determined these being a feature to be worked out during the construction.

### *Stabilization of Orbit and Improvement of Orbit Control*

The magnetic stabilization forces, brought about by the suitable shaping of the guide poles, are known to diminish with an increase in size of the betatron while retaining the same sectional dimension of the acceleration chamber. Wideroe and Touschek have made a theoretical study of the additional stabilization that may be achieved by electron optical lenses. These may, in their simplest form, consist of a set of series connected coils spaced closely around the circumference of the acceleration chamber, the axis of each being tangent to the orbit. The principal advantage of such a lens system is that a greater electron filling of the acceleration chamber should be possible.

The maximum beam current is, in theory, determined by the effect of the mutual repulsion due to Coulomb forces of the electrons. These forces are compensated again by the magnetic forces mentioned above. Excess electrons thus injected are driven out of the beam and end on the walls of the acceleration chamber. The repulsive forces, although decreasing with time after injection of the electrons, are still relatively large. The effectiveness of the electron lens will be greatest in the early stages of the injection process. Wideroe has estimated that, in the 200-Mev design, it is possible to increase the stabilization forces for 10-kv injection about 1350-fold and that this would in turn increase the beam current from 0.1 to 135.0 microamperes. In the 15-Mev betatron the stabilization forces are greater and the relative improvement to be expected is much less. The improvement factor is here about 45 which would increase the current from 0.5 to 22.5 microamperes.

The magnetic lens system was tested out on the 15-Mev betatron at Wrist without any marked success. Further work on this scheme nevertheless is contemplated.

A further subject of concern in the design of the 200-Mev machine was the possibility of energy losses in the machine due to radiation damping set forth by Iwanenko and Pomerantschuk[15] and by others recently.[17,18] Views of Touschek and Lens are that as a result of the radiation damping a radiation of frequency $10^7$–$10^8$ cycles is generated. This radiation, behaving like visible radiation, might in part be reflected by the metallized walls of the acceleration chamber and be conserved. Certainty on this point however did not prevail and it was hoped to study such effects experimentally on a smaller machine before completion of the large 200-Mev betatron. It was also believed that a possible remedy lay in increasing the driving frequency considerably.

## III. BETATRON DEVELOPMENTS OF SIEMENS-REINIGER

Mention has been made previously in this paper of the activities of Dr. Steenbeck and his associates at Siemens-Reiniger. This firm at Erlangen has under development three units two of 5–7-Mev energy and one of 15 Mev. Only the former two are in advanced stages. Of these, the one operating on 550-cycle power is in operation; the other which is to operate on 50 cycles and embody some new ideas in the electron acceleration chamber is still in construction. The 550-cycle unit works quite well and has a radiation output equivalent to that produced by 12–20 grams of radium. The output depends very much on the vacuum conditions and hopeful extrapolation of yields based on the vacuum pressures in use lead to a belief in a possible 50-gram output. The Siemens firm was contemplating putting this machine into production. The betatron is shown in Fig. 9.

---

[17] J. P. Blewett, Phys. Rev. 69, 87 (1946).

[18] L. I. Schiff, Rev. Sci. Inst. 17, 6 (1946).

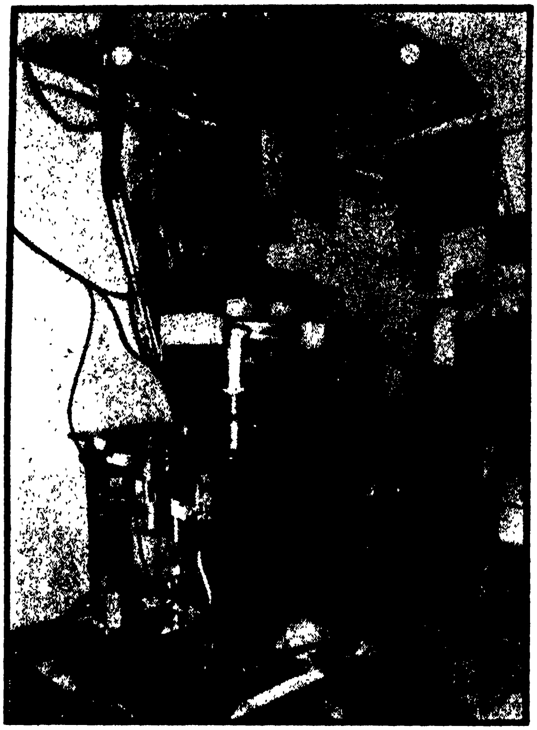


FIG. 9. View of Siemens-Reiniger 6-Mev betatron.

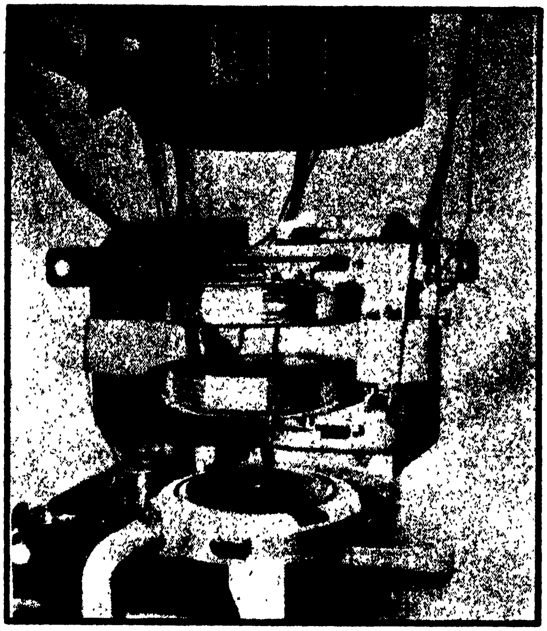


FIG. 10. View of 6-Mev betatron with acceleration chamber removed.

### *General description of 6-Mev, 550-cycle betatron*

This equipment, to be described in more detail below, follows closely the betatron of Kerst in general principles although the constructional details are somewhat different and ingenious. The electrons are injected into the acceleration chamber by a single electron gun which is similar to that of Kerst but modified by the use of auxillary electric deflection electrodes. Radically different methods of electron injection are contained in patent applications but these schemes were mostly in the design stage. A point of departure from the Kerst model lies in the fact that the electrons are injected near the center of the apparatus rather than on the periphery and also that the electrons are injected in that portion of the field between the guide poles where the magnetization forces of stabilization are either weak or absent. After injection into the chamber, the electrons proceed in spiral orbits into a final equilibrium orbit of nearly fixed radius. X-rays are produced after acceleration in the latter orbit by orbit expansion produced by momentary introduction of current into a central winding which increases the flux in the induction pole or may weaken the flux in the guide field.

### *Magnetic Yoke and Windings*

The general features of the yoke, exciter windings, and the acceleration chamber are shown in Figs. 10–13. As shown in Fig. 11 there is a yoke made of bolted stacks of laminations to make an approximately rectangular form with an open center to accommodate the coils and pole pieces. It may be noted that: (a) no gaps are provided for air cooling, (b) no radial construction is used in the large pole pieces which are made approximately circular by the expedient of making parallel lamination packets of different lengths, and (c) that the "staggering" method is employed to improve the general shape of the yoke. Cooling of the main guide poles is largely taken care of by the water cooled exciter windings which are hollow brass tubes on which special Litz wire is wound. Cooling of the injector mount is effected by air passing through ducts parallel to the induction pole and communicating with a suction (or pressure) duct at the base of the machine.

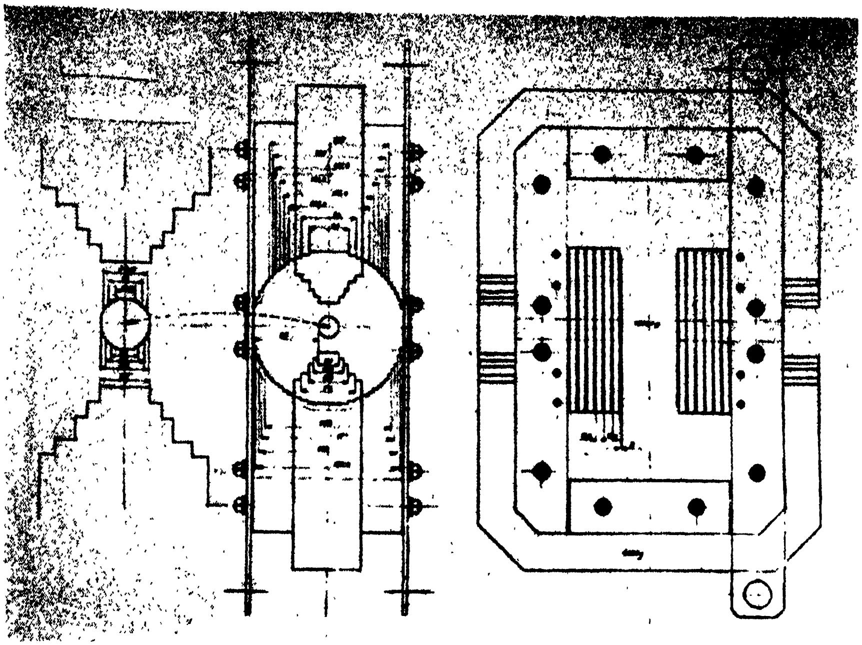


FIG. 11. Photograph of drawing showing yoke and construction of 6-Mev betatron. Since figures are not easily legible, the following scale may be helpful. As printed, reduction is $\frac{1}{16}$.

The guide pole pieces are finished by machining to precise flatness and against these rest the radially laminated packets which are built into the covers of the acceleration chamber and give the guide field its proper shape and intensity distribution. The major portion of the pole piece is thus built integral with the yoke and requires no special arrangement for supporting. At the center of the pole pieces may be seen a hole of approximately circular cross section extending continuously from the bottom to the top of the yoke. It connects at the lower side to the air supply duct, centers the middle section of the acceleration chamber and provides access from the upper side to screws for clamping or tightening the acceleration chamber assembly in place.

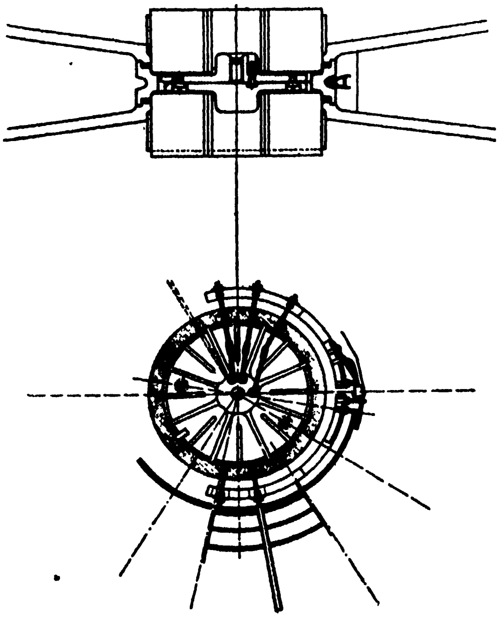


FIG. 12. Electron injection system for Siemens 6-Mev betatron.

The windings consist of 6-mm brass tubing which is wound transversely with five parallel strands of 5 sq. mm total section Litz wire, or of 9 parallel strands of somewhat smaller section Litz wire. These tubing turns were covered with tape, varnished, and formed into coils of 36 turns each (six high six deep) mounted on coil forms which are supported at the sides of the yoke by the binding plates of the yoke.

### *Acceleration Chamber*

The acceleration chamber, shown in Figs. 10 and 12, is the most ingenious and intricate part of the equipment. It is constructed from a porcelain-like material called "Kallait" and made so that it is completely demountable as may be noted in one of the figures. Essentially, it is a body ring with side exhaust port and with circular

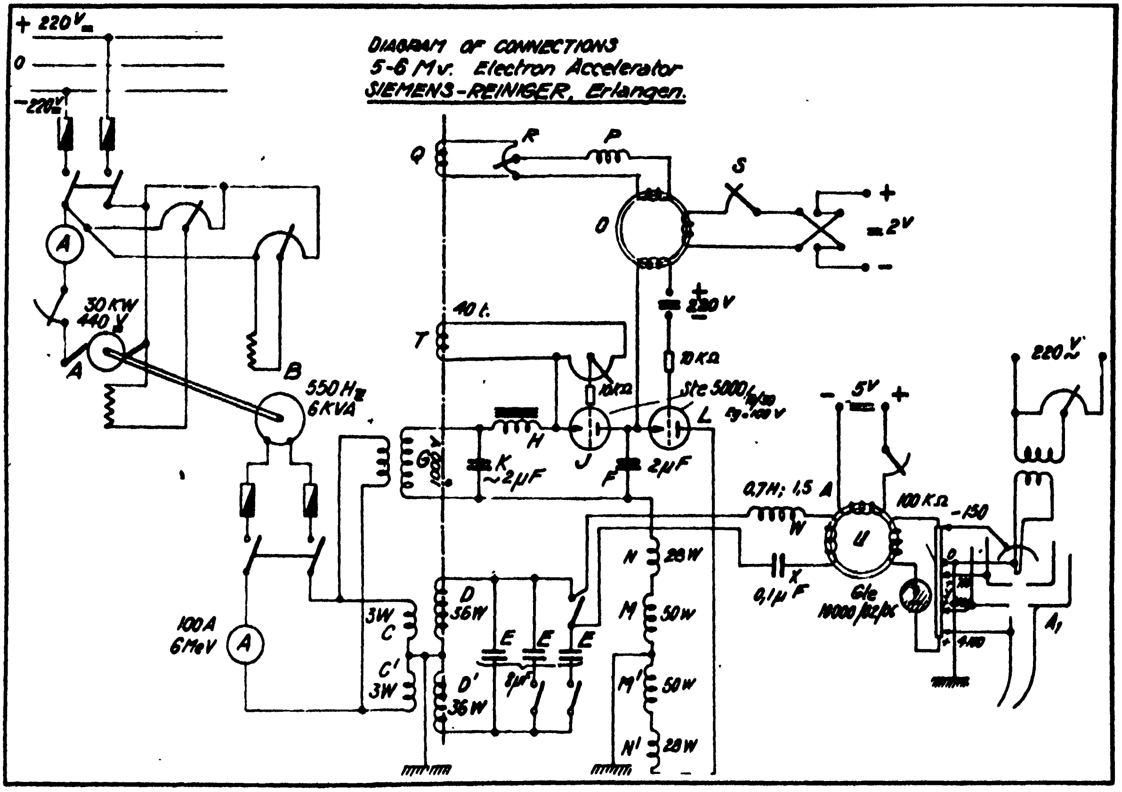


Fig. 13. Diagram of connections, 6-Mev betatron.

grooves for the gasket sealing lids which form the top and bottom of the acceleration chamber. These lids have built into their outward faces the radially laminated iron "faces" which on the outside are flat to match exactly the flat machined guide pole stump and on the inside are shaped to give the proper field shape and intensity distribution to the guide field. Near the center is located a cylindrical well into which is fitted the winding for producing orbit expansion by increasing the induction flux and weakening the guide field. The two lids mentioned have a central opening, the edges of which also make a gasket seat on the central "Kallait" core piece, which furthermore serves for carrying the electron injector structure and also all the necessary electrical lead-ins for its operation. It is cooled as noted above by an air stream passing through the center aperture of the yoke. Another feature of the acceleration chamber is its electrostatic shielding. This is accomplished by metallizing the inner walls of the top and bottom in grid fashion to break up complete electrical circuits about the pole or in its faces. Sections of this metallic coating are connected to radial leads distributed about the periphery of the wall ring. On the outside, these leads are connected together with high resistances and one point is grounded. The body ring also has a metallic spray coating which is contacted by the leads to the coating. On one side of the chamber is the target which is a small rectangular bar of tungsten held with its axis parallel to the field and which is mounted on one of the lead-in stems. On this side of the chamber a slot is provided in the wall of the body ring over which is cemented a thin metal foil window to minimize the absorption of x-rays and electrons passing out from the target. The direction of the radiation beam is approximately 45° to the long axis of the yoke, passing out of the betatron at a point just forward of the pump stem as shown in Fig. 9. This type of construction while apparently complicated is well suited to experimental work especially in view of necessary alterations to test out projected ideas on electron injection and field control.

### *Circuit Connections*

Figure 13 shows the circuit diagram for the 550-cycle betatron. Power is obtained from a 550-cycle motor generator unit consisting of a 440-volt, 30-kw, direct-current motor driving a 6-kva alternator. Frequency is regulated by a manual speed control of the motor and no equipment is used to stabilize the speed. The alternator is a special low voltage machine supplying 100 amperes for 6-Mev operation. This current is passed through six primary turns on the betatron which may be seen passing over the main driving coils in Fig. 10. These are divided into two sets and center grounded, likewise the main windings. Their 36-turn terminals are connected directly to a capacitor bank of 8-mf capacity. A part of the parallel capacitor connection is used to operate a peaking transformer used to provide the voltage pulse for the electron injection. A 5-volt, d.c. winding provides means for shifting the phase of the injection. The output pulse of about 10 kv is passed into a 100,000-ohm potential divider from which various voltages applied to the injector electrodes are obtained. The filament is heated by a separate transformer and is grounded together with one point of the potential divider.

The orbit expansion circuit derives power from a 1-kv utility transformer connected in parallel to the primary coil of the machine. The output passes over a smoothing filter, *K*, and *H*, and a grid-controlled rectifier which controls the charging of the 2-mf capacitor *F*. The grid-control feature of the rectification is not indispensable. Control is obtained from a forty-turn winding *T* divided on the two legs of the machine. A similar winding *Q* supplies power to another peaking transformer *O* having a phase-shifting bias winding. This actuates the thyratron *L* (*L* and *J* are Osram Ste 5000/10/30). The capacitor *T* is discharged at the desired operating point through the two fifty-turn, orbit expansion coils *M* and *M'*. In series with these are two 28-turn, compensating winding *N* and *N'* wound about the legs of the yoke whose purpose is to compensate the magnetization effect of the orbit expansion current pulse in the yoke. It may be noted that the injection and orbit expansion pulses are phased properly by having one peaker transformer actuated by the primary driving current and the other by the voltage derived from the primary flux which is in quadrature therewith.

### *Review of Betatron Patent Ideas of the Siemens-Reiniger and C. H. F. Müller Firms*

While the actual electron accelerator developments of Siemens-Reiniger seem to have covered only patent claims which are fairly well represented in one application[19] entered July 28, 1942, a number of ideas are contained in their present work and in other patent applications which will be reviewed briefly.

In an interesting experimental device use is made of a thermo element on the tungsten target of the betatron to measure the actual electronic current which is converted into radiation and heat. This is a more efficient indicator of output than the electron current passing from chamber to ground.

The use of the positive ion plasma for the purpose of increasing the electron density in the acceleration orbit by virtue of the electronic focusing and stabilization resulting from the positive ion space charge has been considered in a semi-theoretical way. This topic is on the program for further study.

Ideas contained in patent applications include the following:

a. Construction of the magnetic acceleration chamber in such a way that the magnetic guide field poles form the vacuum chamber itself and even contain the magnetizing windings to produce the guide field. This idea has met with experimental difficulties. It was intended for an accelerator for producing fairly strong circulating currents with several hundred kv energies.

b. Use of a built-in saturable magnetic core to keep the guide field constant for a short time during the injection period to increase the available injection time.

c. In relation to (b) use of a constant injection voltage during time in which guide field is constant.

d. Electron injection radially and normal to the field with asymptotic approach to an equilibrium orbit together with mechanism for transfer of orbit to a stable acceleration region or orbit.

e. A ring cathode with auxillary radial electric accelerating field. Applicable only on certain conditions.

f. Polygonal or elliptical ring cathodes to overcome limitations of (e).

g. A 1:2 magnetic field principle operative in asymptotic injection.

---

[19] Siemens Akt. 151,565, VIIIc/21lg.

h. Time dependant guide field further characterized by pre-magnetization of a special magnetic core with a.c. in quadrature with the guide field.

i. Electron source consisting of a pair of oxide emittor plates, above, below, and parallel to median plane of accelerator. An acceleration chamber with an open center.

j. Accelerator as in (i) in which a plurality of electron orbits pass through the axis and produce there a high electron concentration.

k. Acceleration in a stable orbit with either a time variant or time invariant guide fields, together with an electrostatic arrangement for release of an electron beam from the orbit.

l. Orbit expansion by means of local disturbance of guide fields by special windings located at advantageous points or areas in field.

m. Arrangements for d.c. premagnetized guide field, also time variant guide fields.

n. Production of injection potentials by special windings in induction field.

Patent applications[20,21] credited to Dr. Müller of the C. H. F. Müller firm also contain the idea of using a ring cathode and radial electric acceleration fields mentioned in Sections (d)–(f) above. These contain perhaps the novelty of using magnetic shielding material to protect the filament of the cathode from magnetic forces and also have a departure in the proposal: first, to accelerate the electrons by induction within an inner portion of the guide field, then to pass these on by means of an applied radial electric field to an outer region where they may be further accelerated in a stable orbit. In the course of the passage, these are to pass through a retarding radial electric field region, the purpose of which is to bring the radial speed of the accelerated electron back to values that will allow them to be introduced into the equilibrium orbit. Electrons that are slow radially naturally form a closed circular path tending to return them to the cathode. A difficulty presents itself in the delaying field in that it may draw electrons oscillating radially about the equilibrium orbit back into this space. To avoid this difficulty, it is proposed to inject the electron in a line parallel to but above or below the median or orbital plane. This, it is believed, will cause the electron captured by the equilibrium orbit to execute spiral tracks which will make difficult the reentry of electrons into the decelerating radial electric field.

In another Müller's application[22] the idea of spiral orbits about an equilibrium orbit is again set forth, this time as a possible means of avoiding collisions of injected electrons with the injector, the spiraling being so controlled that the electron will miss the injector. This is to be accomplished by applying a circular magnetic field whose lines of force are parallel to the electron orbit itself which field is to be applied only during the beginning of the injection.

In an application[23] it is proposed to release electrons after acceleration in a fixed equilibrium orbit in such a way that, on expansion of the orbit by saturation of the magnetic field, the electrons will move outward into a field varying so rapidly with distance that the spacing between successive turns of the spiral electron orbits becomes quite large. It then will be likely that an electron may strike a target or a window covering a certain angular range in its last revolution. No solution is given as to how the electron is so be properly phased to arrive at the desired instant in the required range.

Another application[24] contains an idea similar to that contained in Part (c) of the Siemen's applications. In this the magnetic guide field or the accelerating field is to be kept constant for a certain short time interval during the electron injection process by introducing into the electrical exciting circuit some controlling rectifier valves. These are usually a pair in parallel, one of which "saturates" in a certain time of operation after application of potential and thus produces the momentary constancy of field required. After a desired time the second grid controlled valve is fired and carries the current as a conductor of low resistance for the remainder of the half cycle. The scheme may be extended to produce a similar effect on the negative half-cycle. A C. H. F. Müller application[25] describes an electron injector in which the equilibrium orbit is allowed to pass through the injector structure without obstruction. The application includes the use of a manifold of such units about the accelerator orbit.

A very late application[26] describes an iron-less

---

[20] C. H. F. Müller M 159,065 VIII 10/21g.

[21] C. H. F. Müller M 159,741 VIII c/21g (43/44).

[22] C. H. F. Müller M 159,714, VIII c/21g (1/3/44).

[23] C. H. F. Müller M 189,257, VIII c/21g (28/12/43).

[24] C. H. F. Müller/Dr. Müller 2000 (date unknown).

[25] C. H. F. Müller Appl. A 3224 of 25/9/44.

[26] C. H. F. Müller, Dr. Mü/Ro/160 Jan. 26, 1945.

betatron. The scheme, in which inductions considerably greater than 10,000 gauss are possible, is to consist of a coil of few turns and an acceleration tube of relatively small dimensions. The coil is of a hemispherical shape to give proper direction to the field lines and proper distribution of the flux through the orbit in the acceleration chamber. Because of the high power requirements due to lack of iron the betatron is to be operated by a sudden discharge of electricity stored in a capacitor bank. It is also proposed to obtain desired circuit time constants by proper design of self-inductance or by inclusion of external circuit impedance. The application further includes possible use of iron in such parts of the equipment where fields of 10,000 gauss are to be used. Iron, however, is not to be used in the regions of most intense fields, that is, 20,000 to 100,000 gauss.

## IV. SUMMARY AND CONCLUSION

The principal betatron developments which have been carried out during the war by several German firms or organizations have been discussed above in some detail. In the way of evaluation of these developments, especially in comparison with developments which have proceeded simultaneously in the U.S.A., the following notes may be added:

The developments seem to have the same basis in preceding scientific literatures; the main reference sources are the works of Wideroe and the patents of Steenbeck and of Kerst, upon which developments agreeing fairly closely in general principle have been made. The success of these developments has been of various degrees. The works of Wideroe and the associated groups MVA and BBC have resulted in an operating 15-Mev electron accelerator described above, and some projected designs. The x-ray output of this unit, reported to be as high as 1 kg of radium equivalent and as low as 30 grams at various times of experimental work, compares fairly well taking the upper figure with early American 20-Mev units allowance being made for difference in maximum energy. The technique of electron injection, although subjected to theoretical investigation, lacks considerably in refinement, perhaps partly due to the difficulties encountered in carrying on experimental work at all. The betatron of Wideroe presents a somewhat different design from that of Kerst in that saturating guide poles are used. This feature, however, has opened up some design possibilities which could not be foreseen in the Kerst design in which the guide fields are produced on the same pole unit carrying the accelerating flux. A comparison in regard to weight indicates perhaps a slightly more efficient use of iron which consideration however is influenced considerably by the dimensions of the air gaps.

The most interesting and remarkable development due to the Wideroe group has been the design of a 200-Mev unit in which a very considerable economy of iron may be found. An interesting comparison may be made in the way of weight, between this design which is to weigh about 40 tons for a 200-Mev rating and the 130 tons of the American General Electric Company 100-Mev machine. In this 200-Mev design certain improvements, such as the use of premagnetization, are considered. At present it may be premature and optimistic to believe that all of these improvements will be practically realized on a large betatron until such have been demonstrated on a smaller scale.

The most successful betatron development in Germany is perhaps that of the Siemens-Reiniger 6-Mev, 550-cycle betatron. This unit has a good output of radiation for its small size and weight and operates with almost unbelievable steady output. The actual details of electron injection follow American practice to some degree with the difference that inside injection is employed. Some interesting differences lies in the electron starting. X-ray production occurs in larger American machines by orbit expansion but somewhat different orbit expansion arrangements are used. The most interesting feature is the adaptability of the unit to various arbitrary injection and acceleration experiments. American betatron practice has advanced to the use of sealed off porcelain acceleration chambers. The use of porcelain, however, was anticipated by Steenbeck who used a tube of such material in 1935.

Although no visit could be made to Berlin, it is understood that the AEG were engaged in betatron development also and that their work is based largely on the preliminary achievements of Wideroe and Steenbeck.

In connection with the developments reviewed above there are to be noted many interesting and ingenious technical details. A review of the patent applications indicates a wealth of ideas which have been systematically studied only in part.

### APPENDIX I

Considering both the force of the electrostatic field of the beam of electrons in the equilibrium orbit, and the magnetic attraction force between this beam and the test electron moving parallel to it at a distance $\Delta r$ their difference will be seen to be

$$\Delta P_1=\frac{eQc^2}{\pi r_0\Delta r}(1-\beta^2),\quad Q\text{: charge in orbit.}$$

The difference in centrifugal and magnetic forces acting on the electron for a small deviation $\Delta r$ from the equilibrium orbit $r_0$ is

$$\Delta P_2=\frac{e^2B^2}{m}(1-n)\frac{\Delta r}{r_0},\quad n\text{: field law index}$$

to a first approximation. Equating these differences and solving for $Q$ the total charge in the equilibrium orbit is

$$Q=\frac{\pi(1-n)(\Delta r)^2(r_0B)^2}{r_0E_0(1-\beta^2)},\quad E_0=m_0c^2/e$$

$$=\frac{\pi(1-n)}{r_0c^2}(\Delta r)^2(E^2+2EE_0)\frac{(E+E_0)^2}{E_0^2}.$$

Eliminating $E$ by means of the damping relation: $-dr/r=\frac{1}{2}dE/E$ from which using $dE=rvdB$ and $dB\sim\omega B_m\cdot 2\pi r/v$

$$E=\frac{2\pi^2r^2\Delta rB_m\cdot f}{-dr}\quad\text{(injection voltage)}$$

$f$: operating frequency

we have finally that the average current supplied $i=f\cdot Q$ is

$$i=\frac{4\pi^3(\Delta r)^3(1-n)f^2rB_m}{-dr\cdot c^2}.$$

# Here and There

### Fiber Society—High-Polymer Joint Meeting

The third meeting of the Division of High-Polymer Physics, which was exclusively devoted to textile and fiber research, was held in Maury Hall of the University of Virginia on September 26–28, 1946. The meeting was sponsored jointly with the Fiber Society. Approximately 105 members of the Division and the Fiber Society and guests registered at the meeting.

The program was arranged by a joint committee under the co-chairmanship of Lewis Larrick and W. A. Sisson. Dr. Larrick also headed the local committee, in which undertaking he was ably and enthusiastically assisted by the staff of the Institute of Textile Technology.

In addition to addresses of welcome by Ivy F. Lewis, Dean, University of Virginia, and Ward Delaney, President, Institute of Textile Technology, three addresses of non-technical nature were given in the course of the three-day meeting. P. Larose of the National Research Council in Ottawa spoke on "Textile Research in Canada." Carl M. Conrad delivered a paper commemorating the life and work of the late Enoch Karrer, Fellow of the American Physical Society and a pioneer in high-polymer research. H. Wickliffe Rose presented, with numerous slides, impressions of Japan and the Japanese textile industry gained on a recent trip there.

Thirteen technical papers were presented. Titles and authors appear hereunder.

W. JAMES LYONS
*Division Secretary*

"Air and Water Vapor Permeability of Fabrics," L. W. Rainard, E. L. Wilson, and A. L. Nestmann, Institute of Textile Technology.

"Moisture Transfer to and from Tire Cords Encased in GR-S," W. James Lyons,* Hilda M. Ziifle, Mary L. Nelson, and Trinidad Mares, Southern Regional Research Laboratory.

"Curl in Woven Textile Fabrics," Ross Whitman, Kendall Mills.

"Tricot," C. W. Bendigo, *Textile World*.

"Tensile and Torsional Properties of Textile Fibers," L. G. Ray, Jr., E. I. du Pont de Nemours and Company.

"The Creep Behavior of Textile Fibers," M. T. O'Shaughnessy,† E. I. du Pont de Nemours and Company.

"Electron Microscopical Studies of Natural Cellulose Fibers," William G. Kinsinger and Charles W. Hock, Hercules Powder Company.

"The Measurement of Irregular Distances and its Use in Studying the Swelling of Cotton Fibers," Enoch Karrer and R. S. Orr, Southern Regional Research Laboratory.

"Performance of Tensile Testing Machines," Herbert F. Schiefer and Josephine M. Blandford, National Bureau of Standards.

"Apparatus for Determining the Coefficient of Friction of Running Yard," Francis B. Breazeale, The American Enka Corporation.

"Spun Nylon," L. L. Larson, E. I. du Pont de Nemours and Company.

"An Investigation of the Load-Deformation Characteristics of Celanese and Wool with Particular Reference to the Use of These Materials in Carpet Pile," R. W. Work, A. M. Thorne, and M. R. Livingston, Celanese Corporation of America.

"The Design of Parachute Shroud Lines as Influenced by Fundamental Energy Absorption Properties," E. R. Kaswell, Fabric Research Laboratories.

---

* Resigned April 1945.
† Now at Brooklyn Polytechnic Institute.

### New Appointments

Three new appointments to the staff of the National Bureau of Standards were recently announced. Dr. B. J. Miller is chief of the recently organized Guided Missile Electronics Section, Dr. H. K. Skramstad is chief of the Guided Missiles Section, and Dr. Ladislaus L. Marton is principal physicist in the Electronics Section.

### Awards

Professor P. W. Bridgman of the Physics Laboratories at Harvard University has been awarded the Nobel prize in physics for 1946.

Dr. Willis R. Whitney, organizer and first director of the General Electric Company's Research Laboratory, has received the Industrial Research Institute Medal for his pioneer work.

### Mathematical Tables Progress

The Mathematical Tables Project, 150 Nassau Street, New York, New York, in its progress report for October stated that work had been completed under the following headings:

Table of spherical Bessel functions, Volume II,
"Manila rope" distribution problem,
Spherical scattering functions for complex arguments,
Computations related to hydraulic analogy of shock wave intersections.

### Information on Preparation of Graphs

The editor of *American Journal of Physics* has announced that anyone wishing to have a reprint of the article "Preparation of Graphs for Physical Papers," which appeared in the March–April 1946 issue of that journal, may receive one without charge by addressing *American Journal of Physics*, Wabash College, Crawfordsville, Indiana. The article suggests "practices intended for the guidance of authors and draftsmen in preparing graphs for reproduction in technical publications." It further states that "most of the recommendations are in accord with those recently made by a committee of the American Standards Association."

### Postdoctoral Fellowship for Women

Announcement will be made early in March of the award of the third Sigma Delta Epsilon Fellowship of $1500. Women with the equivalent of a Ph.D. degree, carrying on research in the mathematical, physical, or biological sciences, who need financial assistance and give evidence of high ability and promise, are eligible. During the term of her appointment the appointee must devote the major part of her time to the approved research project, and not engage in other work for remuneration (unless such work shall have received the written approval of the Board before the award of the fellowship). Application blanks may be obtained from Dr. Louise S. McDowell, 28 Dover Road, Wellesley, Boston 81, Massachusetts. Applications should be submitted before February 1, 1947.

### High-Polymer Lecture Series

During the academic year 1946–1947 a series of lectures on the properties of high polymers is being given at the National Bureau of Standards in Washington. They have been arranged by Robert Simha. The lecturers include the nation's leading scientists in this field from industry and universities. The lectures are open to the public without charge and are held from 7 to 9 o'clock in the evening in Room 214 of the Chemistry Building, National Bureau of Standards. Remaining dates in the course are January 22 and 30, February 27, March 6 and 28, April 24, May 8 and 29, June 5 and 12.

### Necrology

Dr. Wilbur B. Rayton, director of the Scientific Bureau of Bausch and Lomb Optical Company, died October 31. He had been with the company since 1908 and was appointed head of the Bureau in 1926. Two months before his death he had received the Navy Ordnance Development Award.

## Calendar of Meetings

| | |
|---|---|
| **January** | |
| 15–17 | American Society of Civil Engineers, New York, New York |
| 27–30 | American Society of Heating and Ventilating Engineers, New York, New York |
| 27–31 | American Institute of Electrical Engineers, New York, New York |
| 28–30 | Institute of Aeronautical Sciences, New York, New York |
| 30–Feb. 1 | American Physical Society, American Association of Physics Teachers, and Division of High-Polymer Physics (American Physical Society) (Joint Meeting), New York, New York |
| **February** | |
| 15 | American Geophysical Union, Pasadena, California (Section of Hydrology) |
| 16–19 | American Institute of Chemical Engineers, Louisville, Kentucky (Regional Meeting) |
| 20–22 | Optical Society of America, New York, New York |
| 22 | American Mathematical Society, New York, New York |
| 24–25 | Inter-Society Color Council, New York, New York (Joint Meeting with the Technical Association of the Pulp and Paper Industry on the 25th) |
| 24–27 | Technical Association of the Pulp and Paper Industry, New York, New York |
| 24–28 | American Society for Testing Materials, Philadelphia, Pennsylvania |

# Letters to the Editor

## Determination of Magnification in Electron Micrographs

HENRY C. FROULA
*Assistant Physicist, Armour Research Foundation of Illinois Institute of Technology, Chicago, Illinois*
(Received November 1, 1946)

Electron microscopes are often calibrated through the use of a grating replica as a specimen. The assignment of magnifications to subsequent micrographs then depends,

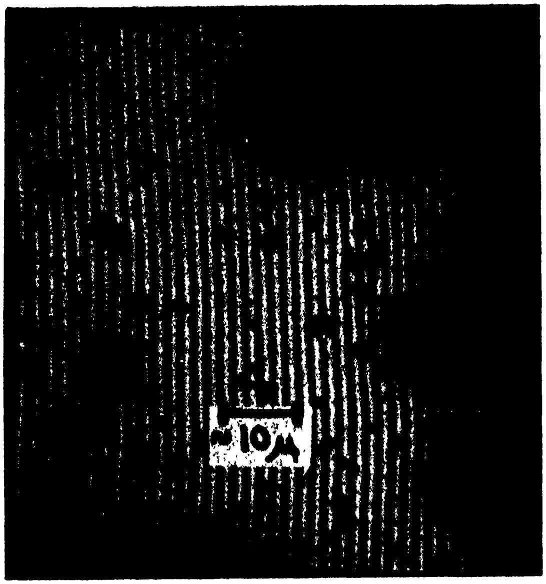


FIG. 1. An electron micrograph of coal dust supported by a 15,000-line per inch grating replica, prepared by the second method described herein. Magnification, approximately 1000×.

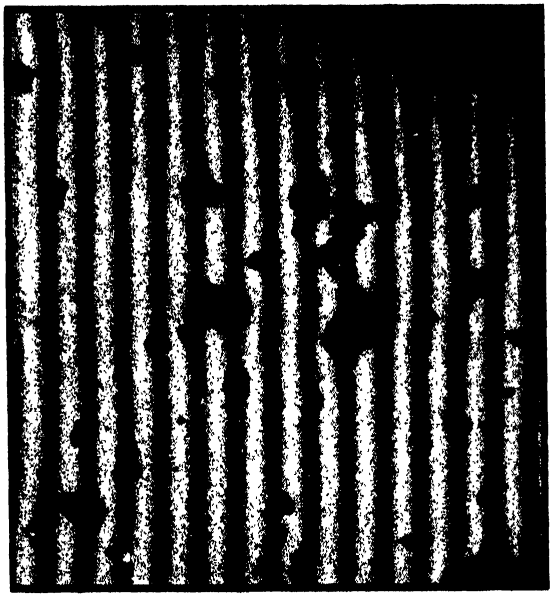


FIG. 2. A limited part of the field of Fig. 1 at a greater magnification, approximately 3000×.

essentially, either upon the constancy of the original calibration or upon a knowledge of how it varies.

Difficulties and skepticism in regard to the translation from the original calibration to subsequent micrographs led to the use of the two techniques described below in two recent studies of small particles.

In one procedure, used in a study of colloidal particles, a regular collodion grating replica was prepared and was mounted on a specimen screen. A small droplet of the sol was then placed on the replica surface. After evaporation of volatile components, the specimen was ready for observation.

In the second method, used in the study of a dry, powder-like material, the particles were mechanically dispersed in an amyl acetate solution of collodion. This mixture was then used in place of the plain collodion solution in making a replica of a grating.

In both of these methods, the replica was prepared in a manner similar to the cellulose-taper-stripping technique described by Schaefer.[1]

In both cases the final image shows the particles with a grating as a background. In this way each micrograph carries its own calibration as seen in Figs. 1 and 2. Furthermore, electron-optical distortions are partially compensated by similar distortions in the grid-like background. However, it will be noted that, while each of these two methods affords a relatively direct means of determining magnification, each still remains subject to inaccuracies caused by such factors as physical distortion of the replica and non-perpendicularity between the electron beam and the replica.

---

[1] V. J. Schaefer, Science 97, 188 (February 19, 1943).

# New Books

## **Matter and Light.** The New Physics.

BY LOUIS DE BROGLIE. Pp. 300, 5×7⅜ in. W. W. Norton and Company, New York, 1946. Price $2.75.

This is an American printing of the translation by W. H. Johnston of Matiére et Lumiére which was originally published in French in 1937. It duplicates the English translation which appeared in 1939.

It consists of a series of twenty-one brilliant essays dealing with modern physics and with philosophical studies raised by these new developments. They are grouped under six headings, as follows: A General Survey of Present-Day Physics, Matter and Electricity, Light and Radiation, Wave Mechanics, Philosophical Studies on Quantum Physics, and Philosophical Studies on Various Subjects. Only two of the essays contain mathematical equations which are, however, familiar to students of modern physics. Since the essays are quite independent of one another, the major portion of the book may be read by anyone who is interested in the revolutionary ideas arising from the development of quantum theory, relativity, and wave mechanics. The line of thought leading to the foundation of wave mechanics by the author is retold in several of the essays with interesting variations.

In this book, the scientist in other fields than physics, the philosopher, and the cultured amateur may find a delightful popularization of the principal ideas of the new physics. The physics student will be pleased by the masterly resumé of modern physics and by the author's insight into related philosophical problems. However, not the least profit to be gained by reading this book is in regard to cultural values, to which the author is keenly sensitive. For example, the final essay, directed to science students, contains the following charge: "the younger generation must undoubtedly acquire learning so that they shall be able to benefit by past experience and by the stock of gathered knowledge; but beyond learning, they must acquire a love of beauty in all its forms, of the art of thinking, and of the expression of thought" and further—"without a zeal to pursue all that is loftiest in the intellectual, the esthetic, and the moral sphere—a civilization, however great the perfection of its material aspects, would soon be no better than a complicated form of barbarism." The intervention of a world war and the invention of the atomic bomb since the first printing of this book has only served to make it more worthy of careful study.

Joseph Valasek
*University of Minnesota*

# New Booklets

*The Ohmite News* for October 1946, published by Ohmite Manufacturing Company, 4835–41 Flournoy Street, Chicago 44, Illinois, features a brief biographical sketch of George Westinghouse.

*Interchemical Review* for summer 1946, published by Interchemical Corporation, 432 West 45 Street, New York 19, New York, has the following table of contents:

High-Speed Microtoming................Mary C. Schuster
What a Research Library Offers..............Lisa G. Otto
Applying Pigment Colors to Textiles.......Winn W. Chase

*The General Radio Experimenter*, published by General Radio Company, 275 Massachusetts Avenue, Boston 39, Massachusetts, features in its October issue an article on "A Peak-Reading Voltmeter for the U-H-F Ranges."

Talk-A-Phone Company, 1512 South Pulaski Road, Chicago 23, Illinois, announced its new 1946 catalog, describing "the world's most advanced and complete line of inter-communication." 12 pages.

Leeds and Northrup Company, 4908 Stenton Avenue, Philadelphia 44, Pennsylvania, publishes a quarterly newspaper-style bulletin entitled *Modern Precision*. Its eight pages contain many short articles and photographs.

Fischer and Porter Company, Department 6F-7, Hatboro, Pennsylvania, has announced its Bulletin 40-A, which describes completely armored variable-area flow rate meters. 8 pages. Available for selected distribution to users of high pressure flow rate instruments.

Battelle Memorial Institute, Columbus, Ohio, recently published a brochure entitled *Research in Action*, a pictorial presentation of the work being done at Battelle in cooperation with industry. 58 pages, many photographs.

Engineers Specialties Division, 980 Ellicott Street, Buffalo 8, New York, has issued a four-page leaflet about its optical projection comparators. The folder illustrates a few examples of complete gauging.

*Bakelite Review* for July 1946, published by Bakelite Corporation, 30 East 42 Street, New York 17, New York, features an article on the use of expanded structural plastics in modern housing. 28 pages, many photographs.

Electro-Voice, Inc., Buchanan, Michigan (formerly South Bend, Indiana), has announced in its *Bulletin No. 131* the new cardyne cardioid dynamic microphone which it manufactures.

The Gaertner Scientific Corporation, 1201 Wrightwood Avenue, Chicago 14, Illinois, has issued its *Description M404*, dealing with the Gaertner M404 Ring Spherometer. 4 pages, free on request.

Dow Corning Corporation, Midland, Michigan, recently published an 8-page pamphlet entitled *DC702 and DC703, Silicone Fluids for Use in High Vacuum Diffusion Pumps*. This is a revision of the bulletin issued on April 15, 1946.

Farrand Optical Company, Bronx Boulevard and East 238 Street, New York 66, New York, has issued a four-page folder announcing its interference filters.

Plastics Divisions of General Electric Company, 1 Plastics Avenue, Pittsfield, Massachusetts, have issued a 14-page booklet entitled *What Are Plastics?* Numerous photographs. Written for the layman. Free on request.

General Radio Company, Boston 39, Massachusetts, recently issued a new 6-page bulletin, *New Variac Continuously Adjustable Transformers*.

Electro-Voice, Inc., 1239 South Bend Avenue, South Bend 24, Indiana, manufacturer of microphones and stands, has published a complete new catalog and selection guide. Free on request.

The Navy's Office of Naval Research, Washington 25, D. C., has republished tables computed by the Mathematical Tables Project of New York for the Massachusetts Institute of Technology Underwater Sound Laboratory of Division 6, National Defense Research Committee. The report is entitled *Scattering and Radiation from Circular Cylinders and Spheres—Tables of Amplitudes and Phase Angles*. The tables were originally computed for a limited audience. The Office of Naval Research believes the report will be of value to scientists engaged in acoustical and electromagnetic wave research, and will therefore make a limited number of these tables available without cost. Scientists wishing to obtain copies of the report should address the Office of Naval Research.

# Contributed Original Research

## Ion Beams in High Voltage Tubes Using Differential Pumping

E. S. LAMAR* AND W. W. BUECHNER
*Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Received July 31, 1946)

Tests on a hydrogen ion gun assembly employing differential pumping are described. Focused beam currents as high as 215 microamperes were realized at the target end of a high voltage tube 6 feet long operating in the range from 300 to 400 kilovolts. Factors which limited the current to 215 microamperes are discussed and suggestions for realizing greater beam currents are presented.

### INTRODUCTION

A NUMBER of investigations of the production and focusing of hydrogen ion beams have been made at the Massachusetts Institute of Technology,[1-5] the ultimate end in view being the use of ion beams together with high voltage vacuum tubes for investigations of the atomic nucleus. In references 4 and 5, which were published together, studies of ion sources and focusing systems were presented which led to a proposed ion gun design using differential pumping. This ion gun, shown in Fig. 1, was built and tested in conjunction with a high voltage vacuum tube. This paper is an account of the tests of the ion gun and vacuum tube assembly which were completed before the recent war interrupted the research, and the conclusions which were reached as a result of these tests. In the sections of this paper which follow, the accounts of the tests of the various parts of the assembly are presented in much the same order in which they were made.

### THE ION SOURCE

The ion source was of the metal capillary type as described in reference 3. The capillary was 3.0 mm in diameter and 3.0 mm long. The outlet was 0.80 mm in diameter. In order to eliminate the need for water cooling, the arc chamber was made from a solid block of copper. The two cathode leads and the one for the anode were brought in through small spark plugs. Both cathode and anode ends of the arc chamber were closed by glass windows to facilitate visual examination of the arc.

The results of the total ion current measurements were in general agreement with those presented in reference 4. They are shown here in Fig. 2. The points shown are the averages obtained from a number of runs, the vertical lines giving the limits of the various measurements. On rare occasions, the output currents were outside the limits shown and approximated those characteristic of the glass ion sources described in reference 4. These rare runs have not been included in Fig. 2. The small fluctuations indicated in Fig. 2 and the larger ones just described are believed to result from changes in surface conditions as was discussed at some length in reference 4. A comparison of the ion currents obtained from the metal source with those to be expected from a glass one is sufficient in itself to show the desirability of completing the engineering development needed to produce a rugged glass ion source. This need is further sub-

---

* At present on leave at the Operations Evaluation Group, U. S. Navy.

[1] E. S. Lamar and O. Luhr, Phys. Rev. 46, 87 (1934).

[2] Lamar, Samson, and Compton, Phys. Rev. 48, 241 (1935).

[3] Lamar, Buechner, and Compton, Phys. Rev. 51, 936 (1937).

[4] Lamar, Buechner, and Van de Graaff, J. App. Phys. 12, 132 (1941).

[5] Buechner, Lamar, and Van de Graaff, J. App. Phys. 12, 141 (1941).

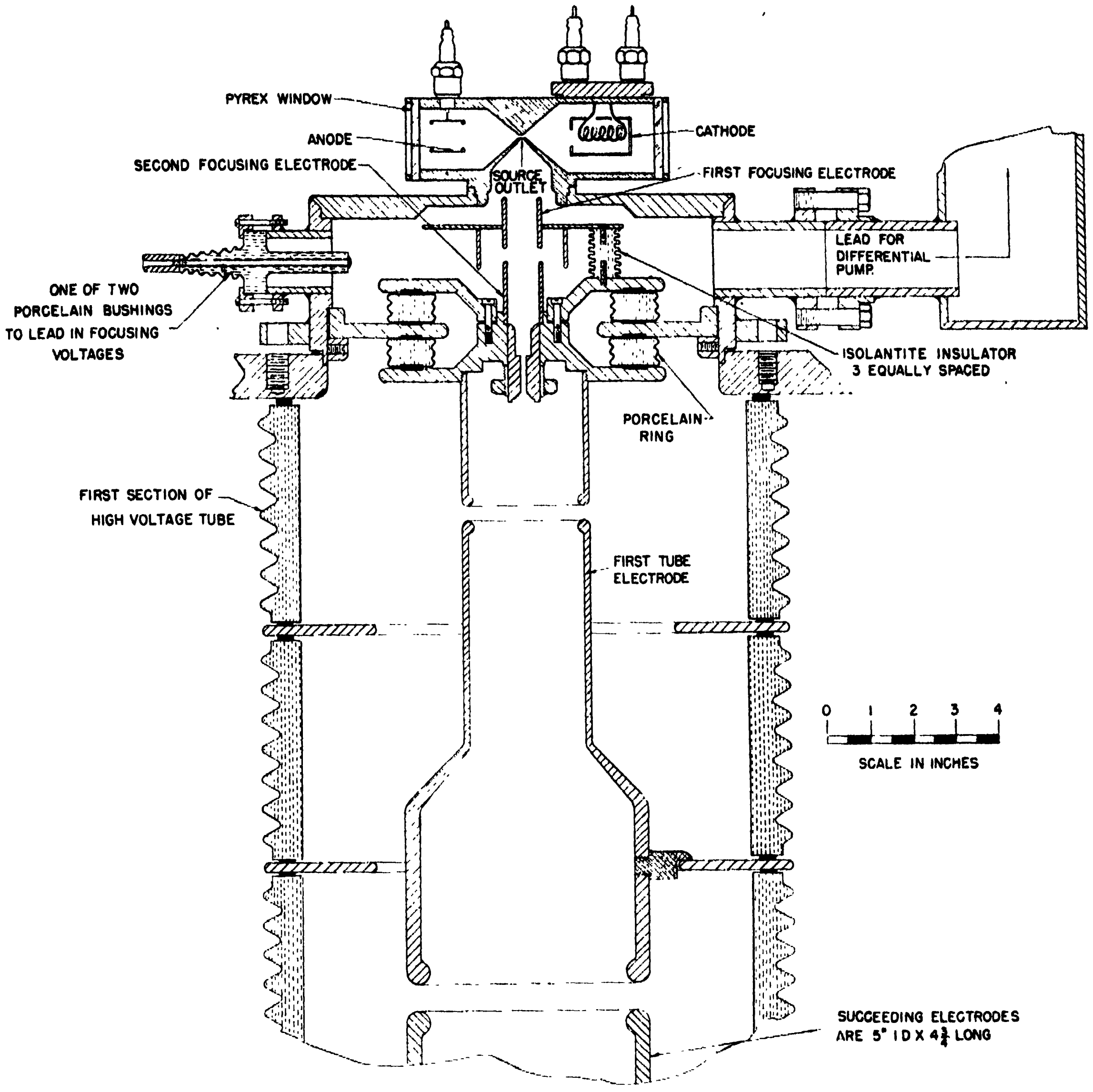

PYREX WINDOW
ANODE
SECOND FOCUSING ELECTRODE
SOURCE OUTLET
CATHODE
FIRST FOCUSING ELECTRODE
LEAD FOR DIFFERENTIAL PUMP
ONE OF TWO PORCELAIN BUSHINGS TO LEAD IN FOCUSING VOLTAGES
ISOLANTITE INSULATOR 3 EQUALLY SPACED
PORCELAIN RING
FIRST SECTION OF HIGH VOLTAGE TUBE
FIRST TUBE ELECTRODE
0 1 2 3 4
SCALE IN INCHES
SUCCEEDING ELECTRODES ARE 5" I D X 4¾ LONG

FIG. 1. Ion gun and initial focusing electrode of accelerating tube.

stantiated by comparing the atomic ion yields of the two types as shown by the mass analysis reported in reference 4.

### THE INITIAL FOCUSING SYSTEM

The initial focusing system, shown in Fig. 1, had exactly the same electrode dimensions as the one described in reference 5. The outlet leading to the high voltage vacuum tube consisted of a canal ⅛ inch in diameter and 1 inch long. In most of the experiments the accelerating potential was applied in two steps: between the focusing cone and first cylinder, and between the first and second cylinders; i.e., a double rather than a single lens was used. With this arrangement, for any given current, greater ion currents through the canal and into the high voltage tube could be realized than with the single lens system. As was discussed in reference 5, for a given total focusing voltage, the aberrations at the source outlet are less for a double lens than for a single one.

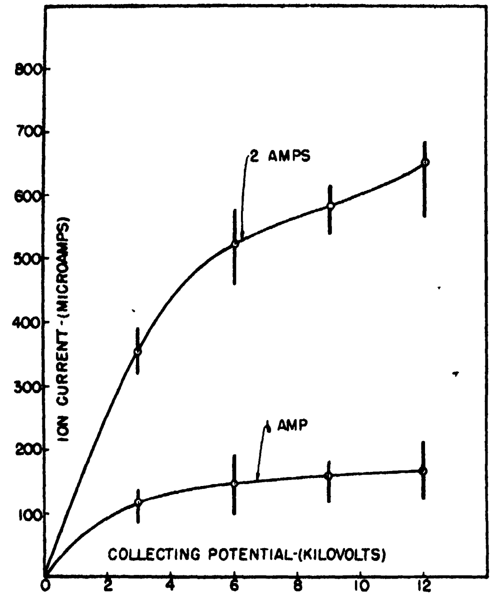


FIG. 2. Output ion current as a function of collecting potential for arc currents of 1 ampere and 2 amperes.

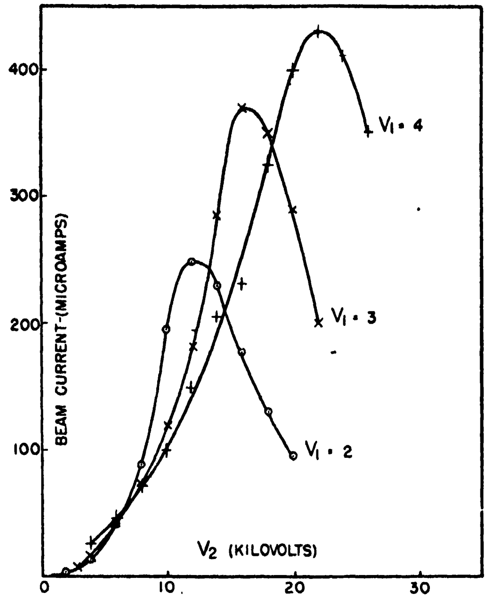


FIG. 3. Ion current focused through differential pumping canal into the high voltage tube as a function of the initial focusing potentials. Arc current = 2 amp.

Figure 3 shows a typical set of measurements of ion current entering the high voltage tube through the differential pumping canal. These measurements were made with all the tube electrodes connected together to form a deep Faraday collector. The potentials $V_1$ and $V_2$ of the first and second cylinders, respectively, were measured with respect to the focusing cone. It is to be noticed in Fig. 3 that for each value of $V_1$, the value of $V_2$ which gives the greatest beam current through the canal and into the high vacuum of the accelerating tube is quite critical. The ratio of voltages for best focus is a function of the first cylinder voltage $V_1$ and of the maximum output beam current. Since the maximum voltage of the power supply for the second cylinder was limited to about 26 kilovolts, it was not possible, in all runs, to read the point of best focus. For those runs in which the peak was reached, the ratio $V_2/V_1$ at the peak is given in Table I as a function of $V_1$ and the maximum beam current. As can be seen from the table, $V_2/V_1$ increases with increasing current and with decreasing voltage. In other words, the greater the space charge in the beam as it enters the second lens, the greater the second lens action required.

### THE PUMPING SYSTEM

The vacuum pump attached to the critical focusing region was of the tapered nozzle type and that for the high vacuum tube was a steel pump of conventional design. Both these pumps used mercury and were provided with dry ice traps. An idea of the net pumping speeds of each of these pumps can be obtained from ratios of pressures in various regions. For the first pump, computing the pumping speed from the ratio of pressures between arc and initial focusing chamber involves a slight complication. The arc

TABLE I. $V_2/V_1$ for best focus.

| Current (milliamps) | $V_1=2$ (kilovolts) | $V_1=3$ (kilovolts) | $V_1=4$ (kilovolts) | $V_1=6$ (kilovolts) |
|---|---|---|---|---|
| 1.57 | | | 5.50 | |
| 0.75 | 6.75 | | | |
| 0.43 | | | 5.50 | |
| 0.37 | | 5.50 | | |
| 0.36 | 6.00 | | | |
| 0.33 | | 5.33 | | |
| 0.31 | | | 5.50 | |
| 0.25 | 6.00 | | | |
| 0.23 | | | | 3.67 |
| 0.20 | | | 4.00 | |
| 0.19 | | 5.00 | | |
| 0.18 | | | 4.88 | |
| 0.16 | 5.25 | | | |
| 0.12 | 4.75 | | | |

pressures at which the measurements were made were such that the mean free path of the gas molecules in the arc was of the order of the outlet diameter. Hence the speed of the outlet was not constant but was a function of the arc pressure. It was found that the pressure ratio between initial focusing region and the arc was

$$R = 5.03(10^{-3}) + 1.88(10^{-2})P_a, \qquad (1)$$

where $R$ is the pressure ratio and $P_a$ is the arc pressure. Using the low pressure part, i.e., $5.03(10^{-3})$, of this ratio, and the speed of the arc outlet, $5.0\times10^{-2}$ liter per second for air, for low pressures, gives 10.0 liters per second for air as the speed of the intermediate pumping system.

The ratio of pressures between intermediate and high vacuum regions was 162. Since, the speed of the ⅛-inch canal was about 0.13 liter per second for air, that of the high vacuum pumping system was about 21 liters per second for air.

### THE HIGH VOLTAGE TUBE AND GENERATOR

The high voltage tube was the one employed by Hill, Buechner, Clark, and Fisk.[6] For these tests the tube was mounted vertically with the source and pumps at the bottom at ground potential. The tube electrodes used by these authors were replaced by similar ones 5 inches in diameter except at the ion gun end. Here the first two electrodes were combined into one. The end of this electrode away from the ion gun was 5 inches in diameter. Several different electrode diameters were tried for the gap adjacent to the ion gun. The one finally adopted was 3″ as is shown in Fig. 1.

At the target end of the tube, several electrode arrangements were employed. The one finally adopted consisted of a deep "Faraday ice pail" having an insulated molybdenum plate at the far end and a two inch diameter opening facing the high voltage tube. A Pyrex window beyond the molybdenum plate made the plate visible from normal eye level. The molybdenum plate was connected to the end plate of the high voltage tube through a bank of B-batteries and a microammeter. The B-batteries provided a retarding potential of 135 volts to prevent secondary electrons from leaving the molybdenum plate.

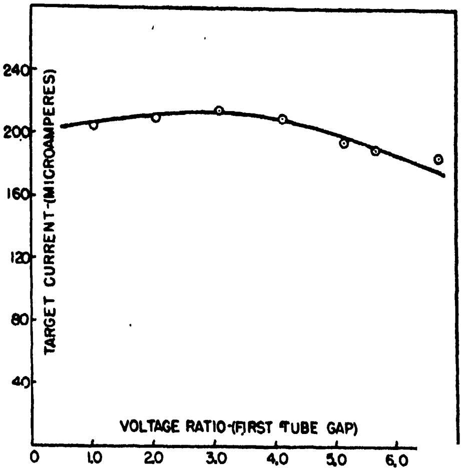


FIG. 4. Target current as a function of the ratio of the potentials across the first gap in the high voltage tube.

The large corona shield which covered the end plate and auxiliary apparatus was made partly of metal gauze to facilitate visual examination of the meter and of the molybdenum plate. The high voltage generator employed was of the belt type.[7] It was suspended from the ceiling and was connected to the corona shield by means of a length of stove pipe.

### TESTS OF HIGH VOLTAGE TUBE AND ION GUN ASSEMBLY

In order to determine the effect of the first high voltage tube gap on the focusing of the ion beam, several different pairs of electrodes were tried. The smallest of these was 1¾ inches outside diameter. Corona control permitted independent adjustment of the potential of the first high voltage tube electrode. Variations of this potential over quite wide limits showed little or no effect on the ion current collected at the target end of the tube. In other words, the first high voltage gap adds little or nothing to the focusing of the ion beam. This is shown in Fig. 4 where the ion current collected at the target end is presented as a function of the ratio of potentials on the two electrodes of the first gap in the high voltage tube.

---

[6] Hill, Buechner, Clark, and Fisk, Phys. Rev. 55, 463 (1939).

[7] We are indebted to Professor J. G. Trump of the Electrical Engineering Department for making this generator available to us for these tests.

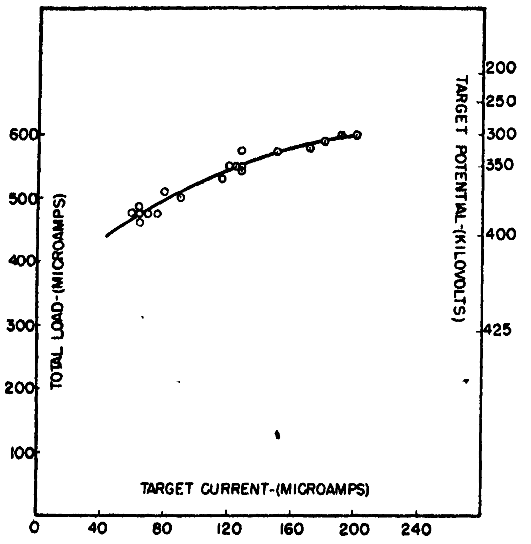


FIG. 5. Total tube current as a function of focused beam current.

The molybdenum plate was employed to give an indication of the state of focus of the ion beam when it reached the target end of the tube. With currents of the order of those presented in Fig. 4, a luminous spot, about $\frac{1}{4}$ inch in diameter, developed in the center of the plate. Because of radial heat conduction from the focal area, the luminous spot size serves only to set an upper limit to the diameter of the focal area. Since this was less than $\frac{1}{4}$ inch, the state of focus at the target end was considered satisfactory.

Attempts were made to realize currents at the target end of the tube greater than the 215-microampere maximum shown in Fig. 4. These attempts were not successful. Whenever the ion gun potentials or the arc current were increased beyond the values corresponding to this maximum, the generator potential dropped very rapidly, the high vacuum tube pressure increased, and electrical breakdown soon followed. These effects indicated the presence in the high voltage tube of a current in addition to that in the ion beam and the increase in pressure indicated that this parasitic current was, in some way, associated with the tube electrodes. The parasitic current may have its origin in the gas, at the electrodes, or both. In this connection, the results of Crenshaw[8] on the loss of energy of hydrogen ions in traversing various gases are of interest. If, under the present conditions, appreciable ionization takes place in the gas, certainly the positive products are not subsequently focused as is indicated by the fact that the target currents are about those to be expected from the primary beam. If the products of ionization strike the electrodes, they may be greatly multiplied by the emission of secondary electrons,[6] and large electron currents may thus originate from the electrodes. Such electron currents may be further augmented by photoelectrons resulting from excitation of the gas along the beam, since it has been observed, with ion beams in glass tubes, that the ion beam is quite luminous even at pressures as low as $10^{-5}$ mm of mercury. There is the further possibility that secondary electrons may be liberated from the electrodes by primary ions scattered from the walls of the canal. This possibility was ruled out in the following way: A new canal, consisting of a diaphragm facing the source and a slightly larger diameter canal facing the high voltage tube, was installed. Although this new canal possessed a much reduced scattering area, the results obtained using it were the same as those with the previous arrangement.

An attempt was made to measure the total current in the high voltage tube, as a means of evaluating the parasitic current. Since this measurement involved certain assumptions, it can be considered as indicative only. First of all, it was assumed that a one to one correspondence existed between generator voltage and generator load. A calibration of the generator was then run using a corona load which could be measured. With this calibration, a measurement of generator voltage provided a measurement of total current. In Fig. 5, this total current is presented as a function of beam current reaching the target. The generator voltage, corresponding to each total load current, is given on the right-hand edge of the diagram. It can be seen from Fig. 5 that, subject to the original assumption, the total load exceeds the beam load by some 400 microamperes.

It is clear that if beam currents of more than a few hundred microamperes are desired, the parasitic current must be reduced. Since the primary cause of this current is collision, either

---

[8] C. M. Crenshaw, Phys. Rev. 62, 54 (1942).

scattering, exciting, or ionizing, in the gas, it is essential that the gas pressure in the high voltage tube be kept as low as possible. If the two pumps of 10 and 20 liters per second, respectively, here employed had been replaced by a single pump used in the conventional manner, a net pumping speed at the source of approximately 1600 liters per second for air would have been required to maintain the same hydrogen pressure in the accelerating tube. Any multiplication of parasitic current by secondary electron emission from the electrodes can be reduced by introducing a diaphragm in the center of each electrode. Since the center is an equipotential plane, no distortion of field should result and the secondary electrons formed off the axis would be stopped after a maximum travel of one tube gap. The focusing results indicate that the hole in the center of the diaphragm could be as small as $\frac{1}{4}$ inch in diameter without cutting off the ion beam. If this were done, it would of course be necessary to arrange the rest of the tube geometry so as to maintain an adequate pumping speed, particularly if the pumping were done from the target end of the tube. Time did not permit investigation of this possibility in the present research. It is, however, strongly recommended for further work.

### ACKNOWLEDGMENTS

It is a pleasure to thank Professor R. J. Van de Graaff for his continued interest in the ion source work and for much helpful discussion. We wish also to express our appreciation to the other members of the high voltage group at the Institute for their suggestions and help in many ways. We are glad to have this opportunity to express our gratitude to the Research Corporation and to the Carnegie Corporation of New York for their generous grants supporting this work.

---

# X-Ray Diffraction Studies of Color Variation of Iron Oxide Pigments

H. Birnbaum, H. Cohen, and S. S. Sidhu
*Cooperative X-Ray Laboratory, University of Pittsburgh, Pittsburgh, Pennsylvania*

(Received August 16, 1946)

Samples of synthetic iron oxide (Goethite) taken at regular intervals from a twelve-day continuous process showed progressive change of color, ranging from pale yellow to dark brown. X-ray diffraction, electron microscope, and optical microscope studies lead to the conclusion that the color change in these samples is mostly caused by particle growth.

THE process of producing synthetic iron oxide pigments has been reviewed by Mattiello.[1] A characteristic feature of these pigments is the progressive change of color from

FIG. 1. X-ray diffraction patterns of samples 1 and 9. Fe$K$ unfiltered radiation, $\lambda = 1.933$A. Diameter of camera 143.2 mm.

---

[1] J. J. Mattiello, *Protective and Decorative Coatings* (John Wiley and Sons, Inc., New York, 1942), Vol. II, p. 300.

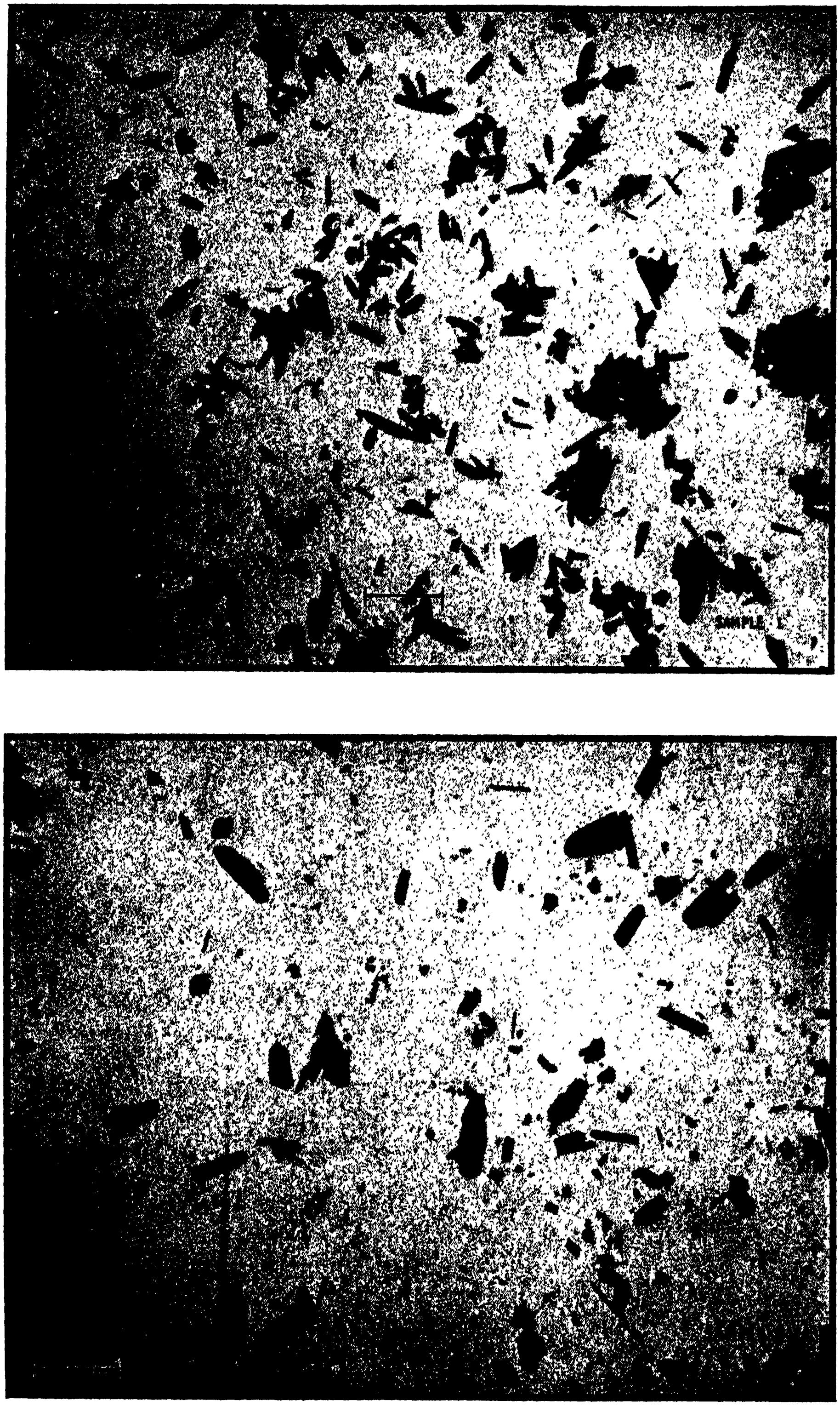


FIG. 2. Electron photomicrographs of samples 1 and 9. Magnification of sample 1: 11250×, and of sample 9, 12000×.

pale yellow to dark brown, as may be observed during processing. The commercially workable methods employ ferrous sulfate, scrap iron, air and steam. According to Fireman[2] a small amount of ferrous sulfate, about 1 percent to 2 percent, occasionally up to 3 percent, is converted into ferric sulfate.

"This suggests that the ferrous hydroxide from the partly hydrolyzed ferrous salt is first attacked by the oxygen of the air, causing the separation of the hydrated ferric oxide and releasing a corresponding amount of acid, which goes to form a certain amount of ferric salt. The latter is then reduced by the iron, which in dissolving restores the ferrous salt. Thus, the new incoming air finds the conditions of the solution exactly the same as the air which preceded it and repeats the same oxidizing action."

Since the chemical analysis of different colored pigments gave approximately the same average composition of 88.0 percent $Fe_2O_3$ and 12 percent $H_2O$, x-ray diffraction and electron microscope studies were undertaken to determine the nature of color variation. The pigments used in these studies were prepared as follows.

## EXPERIMENTAL

To 1100 cc of 9 percent $FeSO_4$ solution were added 200 g NaOH (89.4 percent purity) dissolved in 1000 cc water, under stirring at room temperature. The total volume was brought up to 10 liters and the gelatinous ferrous hydroxide oxidized by slow agitation. The crystallites of ferric oxide monohydrate thus obtained, including mother liquor, were transferred to a tank containing 70 liters of 5.7 percent ferrous sulfate solution and 22 kilograms of scrap iron. As the scrap iron was progressively converted into pigment, more metal was added to maintain a continous process.

The temperature in the reaction tank was kept between 60–65°C during the twelve days of continuous operation. The oxygen, necessary for the reaction, was supplied by forcing considerable air through the solution. Samples, given in Table I, were taken at stated intervals, then filtered, washed, and dried to constant weight at temperature of 125°C. X-ray diffraction patterns of the samples were made by the usual Debye-Scherrer-Hull method employing unfiltered Fe*K* radiation. In addition, all samples were examined with optical microscope for particle size, and electron photomicrographs were made of samples 1 and 9.

TABLE I. Color of pigments, depending on reaction time.

| No. of sample | Color | Reaction time |
|---|---|---|
| 1 | Light lemon yellow | 24 hours |
| 2 | Lemon yellow | 48 |
| 6 | Orange yellow | 168 |
| 8 | Light reddish brown | 216 |
| 9 | Reddish brown | 268 |

## DISCUSSION

X-ray diffraction patterns of samples in Table I were identified as those of the mineral Goethite.[3] Comparison of these patterns showed that the deepening of color in the pigment was accompanied by an increasing sharpness in the diffraction lines. This is illustrated by the photographs of samples one and nine in Fig. 1. Examination of the samples with the optical microscope disclosed a distinct difference in particle size. This was subsequently confirmed by the electron photomicrographs as shown in Fig. 2. While increase of sharpness and greater resolution in x-ray diffraction lines are, in general, an indication of grain growth, or strain free material or both, it appears from x-ray diffraction patterns, electron photomicrographs, and optical microscope studies that the progressive color change of the iron oxide pigments studied here is a visible expression mostly of the particle growth.

## ACKNOWLEDGMENT

The authors want to thank H. C. O'Brien, of the University of Pittsburgh, for the preparation of the electron photomicrographs. The contributions of M. J. Pooley of Hachmeister-Inc., Pittsburgh, Pennsylvania, who assisted in preparing the pigment samples, are also acknowledged.

---

[2] P. Fireman, Ind. Eng. Chem. 18, 286–287 (1926).

[3] C. W. Bunn, J. Sci. Inst. 18, 70–74 (1941).

# Mass Spectrometer for Leak Detection

A. O. Nier,[*] C. M. Stevens,[**] A. Hustrulid,[***] and T. A. Abbott[†]
*The Kellex Corporation, New York, New York*

(Received June 27, 1946)

A simple low resolution mass spectrometer is described which has been used successfully in detecting small leaks in high vacuum equipment. One part of He in 200,000 parts of air can be detected.

## 1. INTRODUCTION

THE problem of finding small leaks in a high vacuum system constructed of metal or glass and metal has always been a difficult and tedious task. Many methods for facilitating the location of leaks have been studied and described.[1,2] Most of the methods depend upon looking for a change of pressure in the system when the part suspected of leaking is waxed or closed with some other substance; or by looking for a change in the nature of the gas or vapor in the system when the part under test is painted with ether, $CCl_4$ and other liquids or when a stream of gas, such as $CO_2$, $H_2$, He, is directed on it. The device for indicating changes in pressure or in the nature of the vapor mixture must be very sensitive in order to find small leaks in systems having many leaks or in systems having surfaces which outgas and hence act as virtual leaks.

Of the various methods for hunting leaks, the scheme of directing a stream of relatively inert test gas at the suspected part of the apparatus and then looking for changes in the concentration or the presence of this particular gas in the system under test has many advantages. Some of these are: (1) gas will not contaminate the system or the pumps; (2) it will enter the system readily through small leaks; (3) there is no danger of closing a leak and having it open some time later; and (4) once a leak is found, there is no time lost in continuing the hunt because the test gas entering the system is quickly pumped out. The ideal "leak detector" is then an instrument which continually samples the gas in the system and instantly notes the presence or change in concentration of the test gas regardless of the presence of other gases or vapors.

A method which has been used with success in the construction of large vacuum equipment employs improved vacuum techniques and a new leak-detecting instrument. The outstanding contributions have been the development of a special mass spectrometer and a method of using it with a tracer gas in the location of minute leaks.[3]

A typical leak-testing arrangement employing the tracer gas method and using a mass spectrometer as a detector is shown in Fig. 1. The system which is being tested is evacuated by the continuous action of the vacuum pumps. An atmosphere of tracer gas may completely surround the system or a small jet of it may be used to probe the suspected zones. Any tracer gas which leaks into the system is rapidly detected by the mass spectrometer which continuously samples the gas being pumped from the system under test. Helium is one of the best tracer gases because it has a low molecular weight, does not occur to any extent in air or in other gases, is not adsorbed on the surfaces, can be readily pumped out of the test equipment, and is easily procured.

## 2. THE MASS SPECTROMETER AS A DETECTOR

The first portable mass spectrometer designed for detecting leaks was built at the University of Minnesota in the spring of 1943. A more recent model was developed in The Kellex Corporation

---

[*] Present address: University of Minnesota, Minneapolis, Minnesota.

[**] Present address: Distillation Products Corporation, Rochester, New York.

[***] Present address: University of Minnesota, Minneapolis, Minnesota. Dr. Hustrulid worked with A. O. Nier at the University of Minnesota on the first sample instruments which were built under OSRD Contract OEMsr-149.

[†] Present address: Standard Oil Company (Indiana), Chicago, Illinois.

[1] J. Strong, *Procedures in Experimental Physics* (Prentice-Hall, Inc., New York, 1942), pp. 134–7.

[2] E. J. Lawton, Rev. Sci. Inst. 11, 134 (1940).

[3] R. B. Jacobs and H. F. Zuhr, J. App. Phys. **18**, 34 (1947).

laboratories and differs from the first model by including an all-metal spectrometer tube and diffusion pump and a special electrode at the collector end of the tube for decreasing the background ion current.

The several components of the mass spectrometer are shown in Fig. 1. The details of the all-metal spectrometer tube are shown in Fig. 2. The gas enters the tube at the top and is ionized by a stream of electrons which are emitted from the heated filament and accelerated toward the trap. The number of ions produced depends on the electron emission and energy and on the gas pressure in the ionizing electron beam.

The ions are pulled out of the electron beam and accelerated by an electric field between the shield and plate containing the source slit. The emission regulator contains the necessary electronic rectifying and regulating circuits for maintaining the desired potentials on the accelerating and focusing plates, as well as holding the electron emission constant at any desired value.

A magnetic field is produced between the iron pole pieces by a permanent magnet. Since the magnetic field is wedge-shaped, it has a refocusing property for divergent ion beams as well as a resolving property.[4-6] The magnetic strength is such that an ion accelerating voltage of about 300 volts is required for collecting helium ions.

After leaving the magnetic field the selected ion beam passes through two baffle plates and a suppressor plate before striking the collector plate. The intensity of the selected ion beam is measured by a conventional battery-powered, negative feedback amplifier with a milliameter as an output meter. As shown in Fig. 2, the first stage of the amplifier which consists of a 954 tube and a $10^{11}$-ohm input resistor is mounted inside the spectrometer tube.

The pressure in the spectrometer tube is measured by an ion gauge whose electron emission is held constant by electronic regulating circuits. The ion gauge is provided with a special filament cut-off circuit consisting of an amplifier tube and a relay. This circuit protects the ion gauge and spectrometer tube filaments from burning out by automatically cutting off their currents when the pressure exceeds a predetermined value.

The tube is evacuated by a high speed oil diffusion pump. A liquid nitrogen trap between the pump and spectrometer tube is used to achieve a residual pressure of $2\times10^{-6}$ mm Hg in the tube. The pumping speed measured at the ionization gauge is approximately 30 liters/sec. This insures a rapid response of the instrument to changes in tracer gas concentration at the throttling valve. In probing, an indication is obtained on the output meter within a few seconds after the probe passes a leak.

### 3. PERFORMANCE OF THE LEAK DETECTING MASS SPECTROMETER

Although the instrument can be used to detect any one of several different gases by the adjustment of the ion accelerating voltage, it is normally tuned to indicate only He$^+$ ions. The selectivity of the instrument is shown by the mass spectrum of air containing a small amount of helium which is plotted in Fig. 3. The amplifier output current is plotted for the different values of the accelerating voltage. The heights of the current

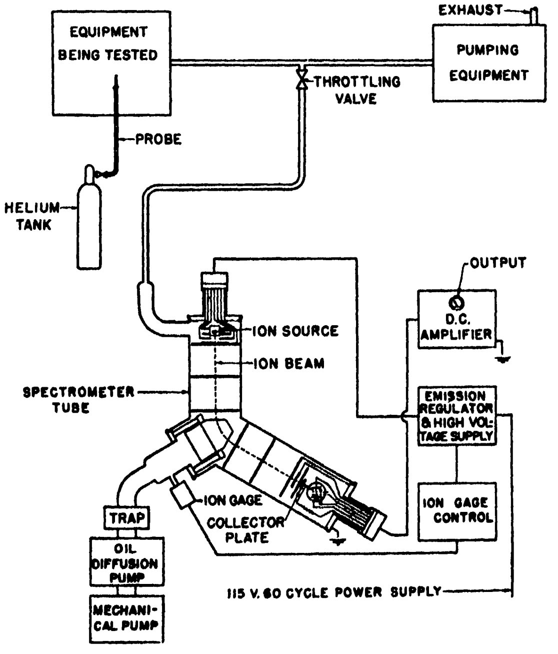


FIG. 1. Schematic diagram of typical leak testing set-up showing interconnections of mass spectrometer components and equipment to be tested.

---

[4] A. O. Nier, Rev. Sci. Inst. 11, 212 (1940).

[5] W. E. Stephens and A. L. Hughes, Phys. Rev. 45, 123 and 513 (1934).

[6] N. F. Barber, Proc. Leeds Phil. Lit. Soc. 2, 427 (1933).

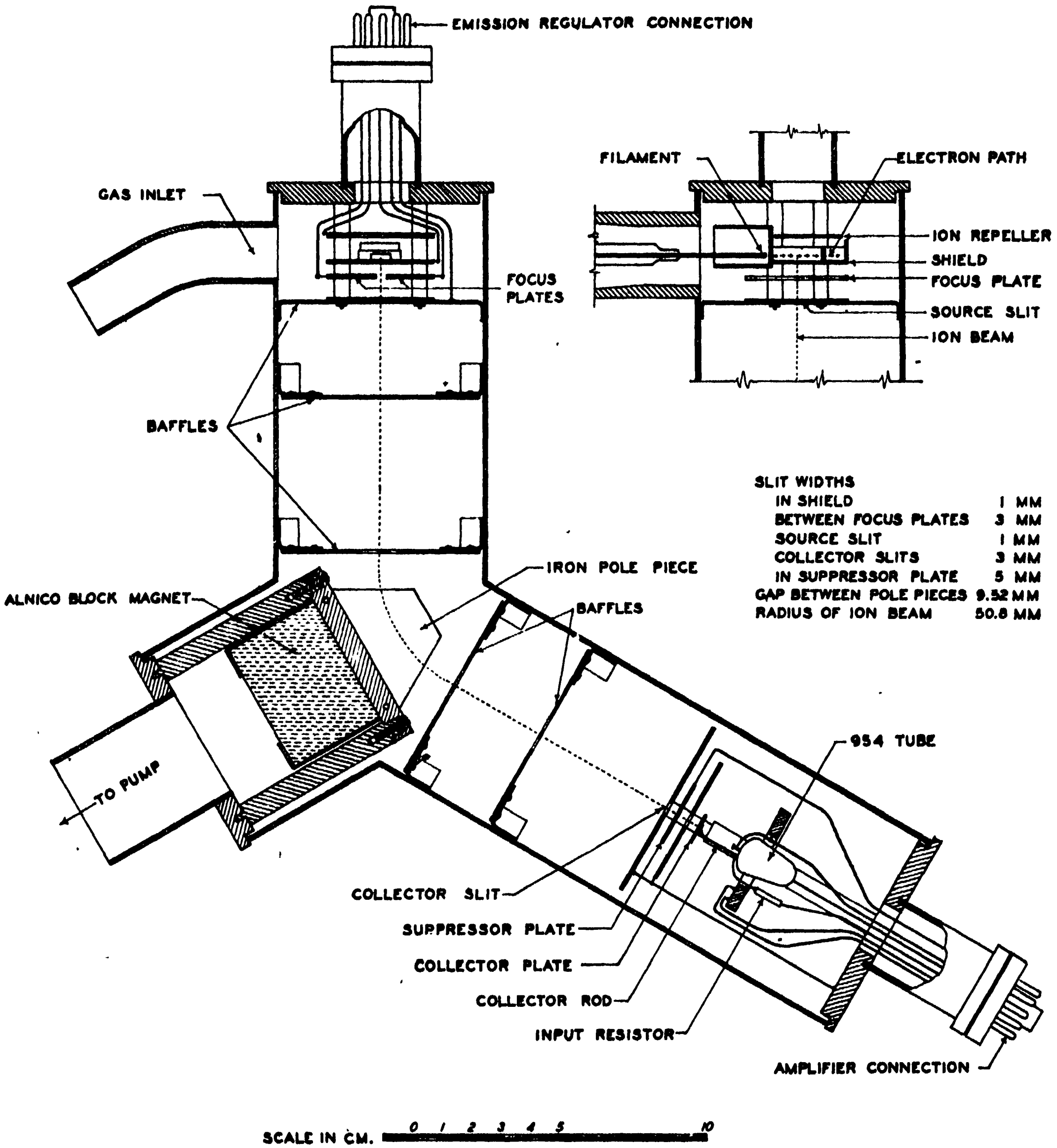


FIG. 2. All-metal spectrometer tube for leak detecting mass spectrometer.

peaks are proportional to the relative concentrations of the ionized gases in the sample. High sensitivity is made possible by two special features of the tube which will be explained below.

If the pressure in the tube is increased, the peaks increase in height since more ions are produced. Unfortunately, increased pressure also greatly increases the width of the peaks at the base especially on the high voltage side of the peak. This produces an increase in background which more than offsets the advantage gained because of the increased height of the peaks. The background is increased because at the higher pressure more nitrogen and oxygen ions lose energy by collision in passing through the region between the source and the magnet and are collected at the same time as are ions of lower mass.

The insertion of a baffle between the ionization chamber and the rest of the tube insures a high pressure in the ion source without a correspondingly high pressure in the rest of the tube. Thus, this baffle materially increases the effective sensitivity of the tube.

A further increase in sensitivity has been obtained by the use of a suppressor plate which is placed just ahead of the collector plate. This plate is operated at a potential near that of the ionizing region so that those ions which have lost energy by collision cannot pass through the slit in the suppressor plate and are not collected on the collector plate. Thus, oxygen and nitrogen ions which have lost energy and would otherwise be collected along with $He^+$ are repelled, and the background is again decreased to materially increase the sensitivity to helium as shown in Fig. 4. Moreover, the $He^+$ ions in passing through the suppressor plate lose almost all their energy so that they produce only a negligible number of undesirable secondary electrons.

Under normal conditions the throttling valve connecting the spectrometer was opened until the ion gauge indicated a pressure of $10^{-5}$ mm in the spectrometer pump lead. This corresponded to an ion source pressure of $2\times10^{-4}$ mm. With an electron emission of 5 ma and a He to air mixture of 1 to 10,000 a $He^+$ ion current of $6\times10^{-12}$ amp. was obtained. This corresponded to a full scale deflection of the output meter. Since it was possible to detect conveniently $He^+$ peaks as small as 5 percent of full scale reading of the output meter, the sensitivity of the instrument was said to be 1 in 200,000.

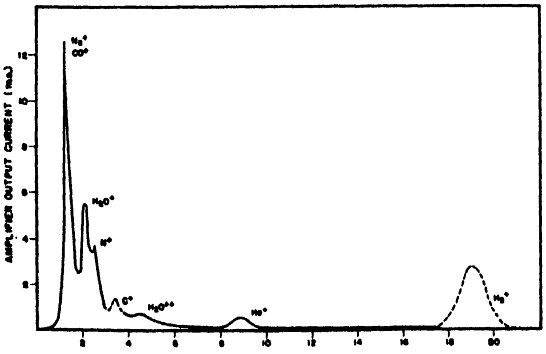


FIG. 3. A typical mass spectrum of air and helium mixture taken on a leak detecting mass spectrometer.

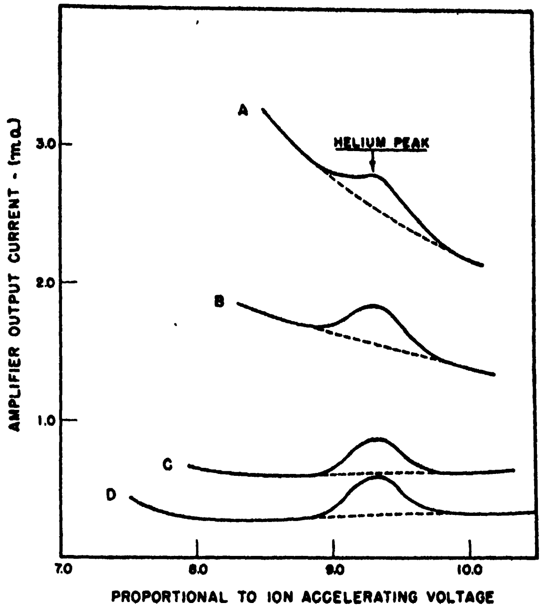


FIG. 4. Effect of suppressor field on mass spectrometer background current in the region of the $He^+$ peak for a sample having one part He to 50,000 parts of air. Pressure in spectrometer tube body $= 1.1\times10^{-6}$ mm Hg; pressure at ion source $= 20\times10^{-6}$ mm Hg.

| Curve | Volts |
|---|---|
| *A* | 0 |
| *B* | +175 |
| *C* | +250 |
| *D* | +400 |

Energy of helium ions approximately 330 volts. Helium ions are collected in case of curve *D* in spite of high suppression voltage due to fact that slit in suppression plate is wide and hence potential of space in slit opening is lower than 400 volts because of proximity of grounded plates on either side.

In order to reduce the sensitivity of the apparatus it was possible to reduce the electron emission or add resistance to the output meter circuit. A sensitivity reduction of a factor of 1000 could be obtained.

## 4. ACKNOWLEDGMENTS

The authors wish to acknowledge the many contributions made by R. B. Thorness to the design and development of the leak detecting spectrometer. This paper is based on work done for the Manhattan Project under Contract No. W-7405 Eng. 23 at the Kellex Corporation and on earlier work at the University of Minnesota under OSRD Contract OEMsr-149.

# New Developments in Vacuum Engineering†

Robert B. Jacobs* and Herbert F. Zuhr**
*The Kellex Corporation, New York, New York*

(Received June 12, 1946)

Methods developed for obtaining vacuum tightness in the K-25 Gaseous Diffusion Plant for the separation of $U^{235}$ are described. A general account of the vacuum engineering of the K-25 Plant also is given. The dynamics of vacuum systems are described, including the use of the mass spectrometer leak detector on rapidly responding systems. New techniques described include, among others, the helium hood method for measuring vacuum tightness and the use of calibrated leaks.

## 1. INTRODUCTION

### 1.1 Degree of Tightness Required

The development of leak detection techniques for the Gaseous Diffusion Plant for the separation of $U^{235}$ was one of the major engineering problems connected with the project. A greater degree of tightness was required than had been achieved, hitherto, in any commercial plant, and the construction schedule did not permit the use of the time-consuming methods generally employed. It was recognized, accordingly, that vastly improved techniques were necessary to greatly increase the sensitivity and speed of testing, otherwise a prolonged period of testing would intervene between the completion of construction and the initiation of operations. After all known methods and several new ones were considered, the mass spectrometer was selected as the basic testing instrument. This paper describes the techniques for leak detection developed with the mass spectrometer, and indicates the general utilization of those techniques.

#### (*a*). *Process Specifications*

Process design required that the inleakage of ambient air into the process stream be no greater than one one-hundredth of a standard cubic foot per hour per thousand cubic feet of plant volume. This is equivalent to a pressure build-up at high vacuum of eight microns (0.00015 lb./sq. in.) per hour. Ambient air must be kept from the process stream, principally because of its water vapor content. Water causes destruction of the uranium hexafluoride, the products of the reaction plugging the minute pores of the barrier. To assure continued, efficient operation of the plant by reducing the effects of inleakage to a minimum, a large portion of the plant equipment was surrounded by air with a dewpoint of −40°F. Elaborate precautions still were necessary to assure that the tightness of the plant was maintained.

Equivalent tightness was required also in the heat exchange equipment, which might otherwise leak coolant into the process system.

#### (*b*). *Precedents*

Industrial precedent for vacuum tightness of the required degree had been limited to compact equipment of relatively small volumes, such as radio tubes, and some of the smaller mercury rectifiers. Indeed, the general trend in industry had been toward accepting leakage in vacuum systems and evacuating continuously with large, high speed diffusion pumps, in order to maintain sufficiently low pressures. The continuously pumped mercury arc rectifier, with a volume of about seventy cubic feet, composed of some fifteen separate parts, had constituted the largest piece of high vacuum equipment generally used on a commercial scale. A few mercury turbine installations had been made with inleakage approximately ten times that permitted the Gaseous Diffusion Plant. Perhaps the modern vacuum still for petroleum oil most closely approximated the complexity and size of the proposed plant, and these units had been considered tight, when the inleakage was approximately 500 times those

---

† This paper is based on work done for the Manhattan Project under Contract No. W-7405 Eng. 23 at the Kellex Corporation.

* Present Address: Distillation Products, Inc., Rochester, New York.

** Present Address: Hydrocarbon Research, Inc., New York, New York.

permitted by process specifications for the Gaseous Diffusion Plant.

### 1.2. Schedule *vs.* Effort Required

Early experience in the laboratories and pilot plants, working in the diffusion process, both here and in England, indicated that the existent testing methods, if applied to a full scale plant would require a testing effort (measured in man-hours) at least equivalent to the effort required to manufacture and assemble such a plant. Indeed, in many laboratories it was common practice, after several men of graduate caliber had devoted a month or two in an unsuccessful attempt to make an experimental set-up tight, to coat vessels with organic "vacuum" waxes. The requirement that the process stream be free from organic contaminants precluded the use of such subterfuge in the Gaseous Diffusion Plant. Further, the exigencies of the project would not permit such an uncertain method of excluding air from the process stream.

Through application of the leak detection techniques developed, using the mass spectrometer as the basic testing instrument, it was possible to improve the process specifications fourfold. A pressure build-up at two microns (0.00004 lb./sq. in.) per hour in the assembled plant was achieved at an effort measuring approximately one percent of the total effort required to manufacture and assemble the process plant.

### 1.3. Pressure Units

Pressure units employed in high vacuum work differ from those most commonly encountered in engineering practices. The unit most frequently used is the micron which is 0.001 mm of mercury absolute or 1/760,000 of standard atmospheric pressure (14.7 lb./sq. in. absolute). Hence, a micron corresponds to a pressure of $1.94\times10^{-5}$ lb./sq. in. absolute. The so-called high vacuum range usually refers to that where the pressure is less than one micron.

### 1.4. Leakage Units

Leakage units are usually microns per hour, when referred to a particular vessel, or more generally, micron cubic feet per hour (mcfh). The first unit only has meaning when referred to a particular vessel. Multiplying the rate of rise of pressure in microns per hour due to inleakage by the volume of the vessel in cubic feet gives the value of the inleakage in micron cubic feet per hour. The latter unit completely defines the leak, since it is independent of the volume of the vessel, and from it the mass rate of inleakage may be computed directly. Hence, a leak of one micron cubic foot per hour means that the leak is of such value as to cause a rate of rise of pressure of one micron per hour in volume of one cubic foot.

## 2. KNOWN TESTING TECHNOLOGY AT THE BEGINNING OF THE PROGRAM

### 2.1. Pressure Methods

It was general practice to test process equipment operating under vacuum, using an internal test medium under pressure, in conjunction with an external indicator for location of the leaks. After all possible leaks were discovered and repaired using this procedure, the system was evacuated and the rate of pressure rise observed as a measure of the inleakage. If the pressure rise exceeded specifications, the system was repressured and the entire cycle repeated. These pressure testing techniques are discussed in more detail in Appendix 7.2. Where applicable to process service lines, these techniques were used in the construction of the Gaseous Diffusion Plant. Design of the process equipment would not permit the application of more than five lb./sq. in. gauge to most of the plant, further decreasing the applicability of pressure testing techniques. However, an improvement of several hundredfold would be necessary for use of this method in the testing program for the plant proper.

### 2.2. Vacuum Methods

Small pieces of vacuum equipment were usually tested using vacuum techniques. Such methods consisted of observing the change in apparent pressure in the system, as measured with either a hot wire or an ionization gauge when the nature of the residual gas in the system was changed by the substitution of a probe gas (or vapor) for the air normally entering through the leak. Most commonly used was the technique of observing a hot wire gauge, while the exterior of the vessel was sprayed with acetone. A more detailed ac-

TABLE I. Analysis of basic leak detection instruments.

| | Mass spectrometer | Differential pirani gauge | Single pirani gauge | Pressure testing |
|---|---|---|---|---|
| Inherent instrument sensitivity (ratio of probe gas to air detectable) | $\frac{1}{100{,}000}$ | $\frac{1}{3000}$ | $\frac{1}{300}$ | Bubble growth at 10 $mm^3$/hr |
| Operating pressure | <0.1 microns | <25 microns | <50 microns | 45 psig |
| Speed of probing | 5 ft./min. | As fast as area can be wet with spray gun | As fast as area can be wet with spray gun | Observation of smooth soap film for at least 5 min. |
| Smallest leak detectable on vessels having a pressure build-up at 200 microns per hour and a volume of: | | | | |
| <1 cubic foot | <0.01 mcfh | 0.10 mcfh | 1.0 mcfh | 5 mcfh |
| 1–10 cubic feet | 0.01–0.02 | 0.10– 1.0 | 1.0– 10.0 | 5 |
| 10–100 cubic feet | 0.02–0.2 | 1.0 –10.0 | 10.0– 50 | 5 |
| 100–1000 cubic feet | 0.2 –2.0 | 10.0 –50 | 50 –500 | 5 |

count of these methods is given in Appendix 7.3. Results obtained using these methods indicated that some improvement over pressure testing techniques could be reasonably expected. However, a corresponding increase in the caliber of the necessary personnel using these methods was also indicated, and the increase in the speed of testing was uncertain.

### 3. DEVELOPMENT OF DYNAMIC VACUUM TESTING TECHNIQUES

Study of the evidence on hand indicated that further advances in pressure testing methods would, at best, advance their useful sensitivity to equivalence with the then existent vacuum techniques, at the expense of an increase in testing time. It was evident that improvement of the existing vacuum techniques could have been effected, but again only at the expense of additional testing time. The increased time factor made both proposals most unattractive. However, new vacuum techniques were visualized which would permit a tenfold improvement in sensitivity and a simultaneous improvement in the speed of testing. These new techniques included:

1. Use of selective instruments; that is, instruments which give a nearly null reading for air and residual gases, responding only to a probe gas.
2. Use of these instruments dynamically; that is, as adjuncts to a high speed evacuating system permitting their use under optimum conditions.
3. Use of the selectivity of the indicating instruments to permit estimation of the amount of leakage even in the presence of outgassing and unrepaired leaks.

It was therefore decided to investigate and develop, if practical, a selective, continuously sampling leak detector, and to determine the optimum conditions for its use.

### 3.1. Leak Detectors Developed and Investigated

The Kellex Vacuum Engineering Group developed the Differential Pirani Gauge as a new continuous sampling unit, and investigated the use of the optical spectrometer as a continuous sampling unit. The University of Minnesota group had developed a compact, specialized mass spectrometer.[1] The use of the instrument, as a continuous sampling (or dynamic) instrument, was developed by the Kellex group. All three of these instruments were selective. Comparison of the Differential Pirani Gauge and mass spectrometer to older methods are made in Table I.

#### (a). *Optical Spectrometer*

The optical spectrometer, off-hand, would appear to be a suitable instrument for leak detection work. It is simple, its response is instantaneous, and its sensitivity for certain probe gases is very high. However, under actual working conditions, it is found to be unreliable due to the uncertain "clean-up" of the discharge tube used as a source after exposure to different gases.

#### (b). *Differential Pirani Gauge*

This instrument consists of two hot wire vacuum gauges known as Pirani Gauges. These

---

[1] A. O. Nier, C. M. Stevens, A. Hustralid, and T. A. Abbott, "Mass spectrometer for leak detection," Contribution from the Kellex Corporation, Research Paper C3.

are connected electrically in opposite arms of a Wheatstone bridge. The two gauges are connected to the vacuum line of the system under test by a short "tee" connection. A cold trap is placed on one of the arms of the tee leading to the gauges, but not on the other.

When air is the only gas present, the pressures in the two gauges are equal and the galvanometer reads zero. However, when a leak is probed with a liquid such as methanol, the latter reaches one gauge only, due to the presence of the trap in front of the other, and the Wheatstone bridge is thrown off balance. The instrument thus acts as a detector for condensible probe gases, being little affected by the presence of other gases.

### (*c*). *Mass Spectrometer*

The mass spectrometer has been used by physicists for many years, principally to determine the isotopic constitution of different substances. In its usual form it is a large and rather complex instrument requiring a highly skilled operator.

A simplified instrument specifically designed for vacuum testing was developed at the University of Minnesota. A considerable reduction in size and weight was made, making a portable instrument possible. Several electronic controls were incorporated reducing the number of manual adjustments required. However, carefully trained operating and maintenance crews are required to use these instruments successfully for leak detection on a large scale.

In popular literature, the mass spectrometer is referred to as an "Atom-Sifter" due to the ability of the instrument to "sift out" and measure the concentration of atoms of any kind.

This characteristic of the mass spectrometer was the basis in choosing a probe medium. This medium should have the following properties:

(1) It should be unique in the mass spectrum of residual gas in the system under test and practically non-existent in the normal atmosphere surrounding the equipment under test.
(2) It should be readily removed from the system by pumping and should not contaminate the system.
(3) It should be of low molecular weight and low viscosity, and should be readily available.

Helium fulfills all of the requirements and was chosen as the probe gas for use in the testing program.

The instrument works in the following manner: Referring to Fig. 1, the gaseous mixture to be analyzed is pumped into that part of the spectrometer which includes the source. When used as a leak detector, the spectrometer is fed through a throttle valve connected to the main vacuum lines. The pressure in the spectrometer is kept less than $10^{-4}$ mm by continuous pumping. In the source of the spectrometer, molecules of the different elements present are ionized by an electron beam. The number of ions of any particular element so produced in a function of the concentration of that element in the gaseous mixture fed into the spectrometer.

The ions thus formed are collimated and given a certain momentum toward the analyzer by suitable electrical potentials. On passing through the wedge-shaped magnetic field, the ions are deflected through certain angles which depend upon their respective masses.

In Fig. 1, the electrical fields are adjusted so that $He^+$ ions enter the collector where they are measured. Ions of greater or of smaller mass fall on either side of the collector slit and are not measured. Ions of any given mass may be brought

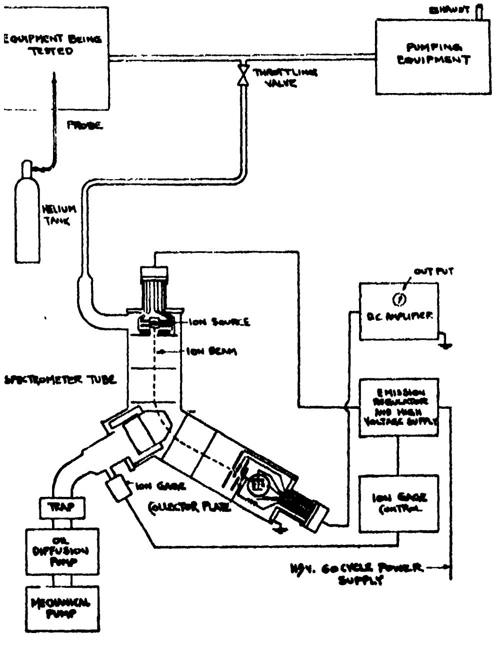


FIG. 1. Schematic line diagram showing use of mass spectrometer leak detector.

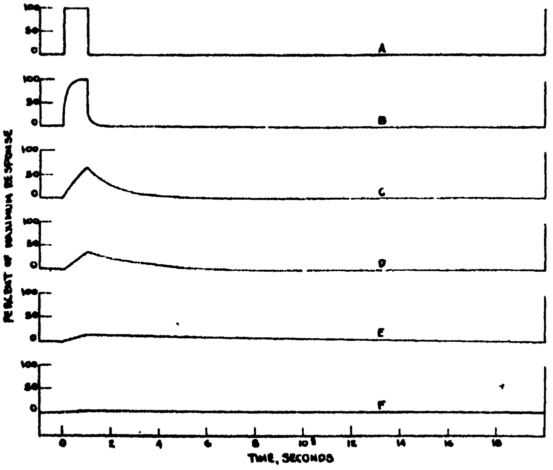


FIG. 2. Appearance and disappearance of probe signal for systems of varying response characteristics. (Probing is for a period of one second beginning at $T=0$.) Curve A is the probe pulse at the leak. Curves B through F have $S/V$ values of 300, 60, 30, 6, and 3 minutes$^{-1}$, respectively, when $S$ is expressed in cfm and $V$ as ft.$^3$.

into the collector by properly adjusting the electrical fields. For leak detecting purposes, the instrument is adjusted for He$^+$ ions, since helium is a most satisfactory probe gas. Theoretically, the instrument, when thus adjusted, will show a zero reading, unless helium is present in the gas stream under analysis. Actually, however, due to necessary physical imperfections in the apparatus, there is a continuous "background" reading. The presence of helium then is indicated by an increase over the normal background reading.

A separate report[1] gives detailed information on the construction and operation of the instrument developed at the University of Minnesota. Accordingly, we are concerned here, principally, with the performance of the instrument as a leak detector.

The first mass spectrometers used for leak testing purposes were all glass instruments using mercury diffusion pumps. The fragility of these machines for plant use was recognized, but the magnitude of the trouble caused, thereby, was discovered only after attempting to institute the production testing program. The entire spectrometer was redesigned by the Kellex Special Instrument Group. In addition to greater durability due to the practically all-metal construction (only the filament stem and the ion gauges remained glass), the new tube possessed a higher inherent sensitivity from design, and a greater sampling capacity through the substitution of an oil diffusion pump for the mercury model.

### 3.2. Dynamic Requirements for Rapid, High Sensitivity Vacuum Testing

Use of a high sensitivity, dynamic, leak detector such as the mass spectrometer is of little avail if the system under test does not possess the required characteristics. It is shown mathematically, in Appendix 7.4, that the ratio of $S/V$ (pumping speed of the system with respect to the volume pumped) is of greatest importance, both with respect to dynamic sensitivity and speed of leak hunting. In Fig. 2, the response (at the leak detector) after exposing the leak to probe gas for one second is plotted for different pumping speeds.

The sharpness of the response to probing will, in a large measure, determine the ability of *any leak detector* to function efficiently. Figure 2 illustrates the impossibility of attaining a sharp response without a sufficiently high $S/V$ ratio: i.e., about 6 or more reciprocal minutes.

When a large leak is probed, the system becomes temporarily "flooded" with a probe gas, and it is impossible to continue leak hunting until this gas is removed. In Fig. 2 probing is discontinued at the end of 1 second. Those parts of the curves for time greater than 1 second represent the "clean-up" period. The length of time required for "clean-up" is a function of the ratio $S/V$. Thus, for example, if a large leak is probed until a probe response of 25 percent of maximum is obtained, and pumping is then continued until all but 0.5 percent of the probe gas is removed, the "clean-up" times indicated in Table II are obtained. These calculations have been confirmed experimentally.

Loss of time by slow clean-up can prolong the leak hunting period considerably, perhaps much more than is indicated in Table II. The reason for this is the following: When one is hunting for leaks with too small a pump (low $S/V$), the dynamic sensitivity is low, and as a result, leaks

TABLE II. Evaluation of clean-up time as function of $S/V$.

| $S/V$, minutes$^{-1}$ | 300 | 60 | 30 | 6 | 3 |
|---|---|---|---|---|---|
| Clean-up time, seconds | 1.7 | 5.8 | 11.1 | 139 | 546 |

must be carefully and repeatedly probed to establish their exact location. Consequently, by the time this is done, a considerable quantity of probe material will be introduced into the system. Since this must be removed before probing can continue, the leak detecting crew finds it necessary to remove a large quantity of probe gas at a slow rate, which is doubly time consuming.

A simple experiment was made which confirmed the above predicted results. The calculations have since been confirmed many times in practice, both under mass production and field conditions.

It must be recognized, that the *S*/*V* ratio refers to the speed of the entire vacuum system, including the process equipment under test. Thus the largest high speed vacuum pump will be of little use if the rest of the system throttles the flow to the pump. Rather, the use of a pump oversize for the system speed will further penalize the sensitivity, by making it impossible to obtain an adequate sample in the gas evacuation line through the leak detector. It is necessary that the lines of the system under test be sized to permit adequate speed of pumping. This can be done by sizing them for molecular flow according to the formula: Speed $= 14D^3/L$, where the speed is expressed in cubic feet per minute, $D$ is the internal pipe diameter in inches, and $L$ the length in feet. One of the more important functions of the Vacuum Engineering Division was to check all design to assure that the sufficient speed was always attained at the test connections.

The need for high speed pumps is apparent from the preceding discussion. These pumps had other special requirements:

1. The pumps and all associated vacuum equipment, i.e., valves, gauges, and piping, had to be tight, and engineered to maintain this tightness.
2. Provision to reduce the back diffusion of diffusion pump oil, without seriously impeding the speed characteristics of the pump, was necessary to prevent the contamination of the process equipment under test.
3. The complete pump unit, including mechanical and diffusion pumps, and all necessary controls, had to be self-contained and portable, and should require only connections to the system under test, water, power, and drain lines, to permit operation.

A typical vacuum pump unit is shown in Fig. 3. For initial evacuation, provision is made to by-pass the diffusion pumps through a roughing line. Pump units of suitable design were developed by Westinghouse Electric Corporation, National Research Corporation, and Distillation Products, Incorporated, in conjunction with the Kellex Vacuum Engineering Division.

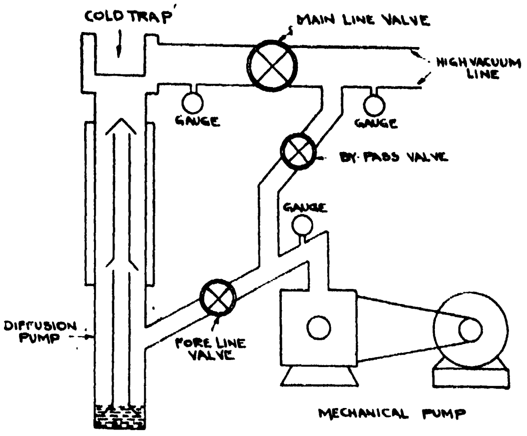


FIG. 3. Typical pumping unit.

## 4. MEASUREMENT OF TIGHTNESS

The general plan of attack, to secure over-all plant tightness, was to apportion the total allowable leakage among the different plant components, and to check individually for tightness each such component, one or more times prior to its installation.

This measurement of tightness, involving about 1,000,000 individual tests in all, threatened to be a serious bottleneck in the fabrication schedule. The only known method for measuring tightness consisted of evacuating the equipment sufficiently long to outgas it thoroughly and then to measure the pressure rise. This method usually required an average of 12 hours of continuous evacuation for each piece of equipment tested. Since the estimated number of vacuum tests on the individual components of the plant totaled over 600,000, it is evident that approximately 1000 leak rate stations each working more than 8000 hours would be required to prove the tightness of equipment before installation. This estimate does not include the time spent leak hunting, setting up the equipment for test, and retesting of rejects.

Evaluation of the mass spectrometer as the basic leak detection instrument indicated that its

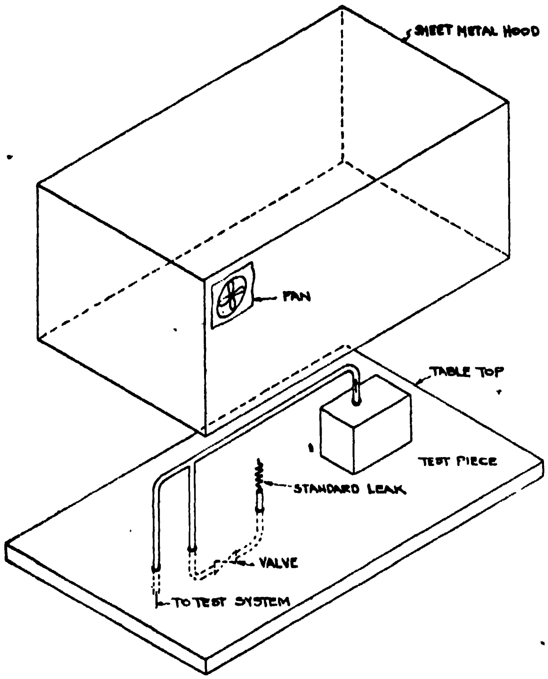


FIG. 4. Typical hood test set-up.

quantitative aspects could be adapted to measurement of the inleakage into the vacuum system. Accordingly, the so-called "Hood Method" for tightness testing was devised. Using this method in conjunction with the rapidly responding dynamic system for leak detection, the testing load for production testing was reduced to approximately 100 stations each working 2450 hours.

The Hood Method consisted of surrounding the equipment under test with a gas-tight hood in which a fixed concentration of probe gas was maintained, and comparing the amount of the probe gas entering the system through the leaks in the equipment to the amount of probe gas entering through a standard calibrated leak. This comparison was made by direct reading of the output of the mass spectrometer. The results thus obtained have been quantitatively checked against results obtained by measuring the inleakage of thoroughly outgassed vessels by the usual pressure build-up method.

A typical hood set-up is shown in Fig. 4. For small objects several manifolds are attached to one pump-leak detector combination, allowing the objects to be set up or taken down on one manifold, while the units on the other are under test. For larger equipment, individual pump-leak detector combinations are required. The test is made in the following manner:

With the valve leading to the calibrated leak in an open position helium is forced into the hood until its concentration there is sufficient to give a predetermined reading on the leak detector. This reading (reading 1) is proportional to the total inleakage consisting of: (a) unknown vessel leakage, (b) known leakage from the calibrated leak.

The valve leading to the calibrated leak is then closed, and a second reading taken (reading 2). This reading is, of course, proportional to the unknown leakage alone. From these two readings the following equation may be derived:

Unknown leakage =

$$\frac{\text{Rdg }(2)}{\text{Rdg }(1)-\text{Rdg }(2)}\times\text{calibrated leak.}$$

This equation gives the unknown leakage in terms of the two readings referred to above and the value of the calibrated leak. Knowledge of the helium concentration in the hood is not necessary. The only requirement is that this concentration remains substantially constant during the time the readings are taken. Normally, this requires from five to ten minutes. Well constructed hoods show practically no decrease in helium concentration for considerably longer periods.

A modification of the hood technique, that has proved extremely useful for testing "hidden" leaks, involves the use of undiluted helium on both the standard leak and the unit under test. This method is best achieved by evacuation of the re-entrant interior volume suspected of

TABLE III. Contribution of various process components to inleakage (based on manufacturer's specifications).

| Component | Volume contribution per $100\ \text{ft.}^3$ | Inleakage specification pressure build-up, microns/hr. | Inleakage contribution, mcfh |
|---|---|---|---|
| Diffusers | 68.6 | 0.25 | 17.1 |
| Centrifugal pumps | 2.2 | 0.50 | 1.1 |
| Process drums | 3.8 | 0.1–1.0 | 0.4 |
| Process piping and valves | 25.4 | 1.00 | 25.4 |
| Instruments | <0.1 | <1.00 | <0.1 |
| Total per 100 cubic feet |  | 0.44 | 44.0 |

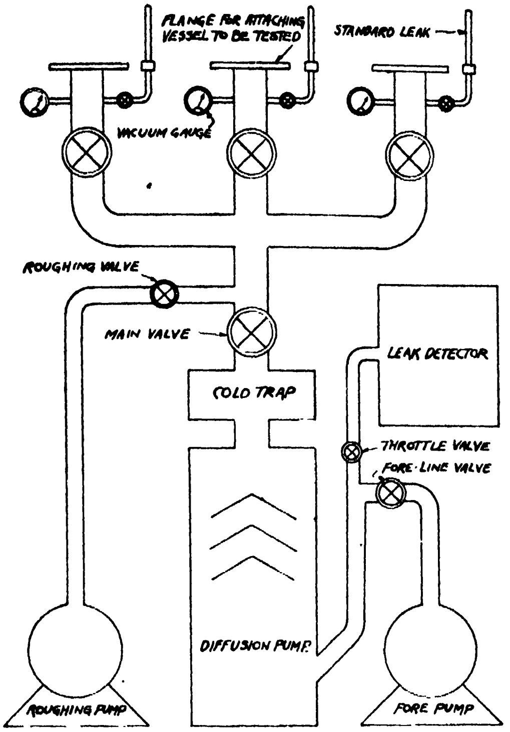


FIG. 5. Manifold for medium volume—medium production vacuum testing.

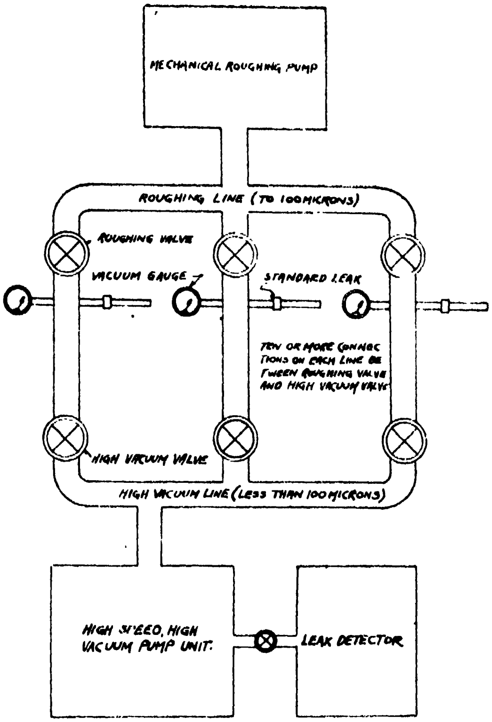

MECHANICAL ROUGHING PUMP
ROUGHING LINE (TO 100MICRONS)
ROUGHING VALVE
VACUUM GAUGE
STANDARD LEAK
TEN OR MORE CONNEC TIONS ON EACH LINE BE TWEEN ROUGHING VALVE AND HIGH VACUUM VALVE
HIGH VACUUM VALVE
HIGH VACUUM LINE (LESS THAN 100MICRONS)
HIGH SPEED, HIGH VACUUM PUMP UNIT.
LEAK DETECTOR

FIG. 6. Manifold for small volume—high production vacuum testing.

leaking into the vacuum space proper, and breaking the vacuum with pure helium. Then the reading of the mass spectrometer output is compared to that obtained when a standard leak is surrounded by pure helium.

### 5. PRODUCTION TESTING

Experimental evidence indicated that the process specifications could probably be reduced from eight microns per hour to two microns an hour through the use of the methods just described. It was evident that most careful preassembly testing would be necessary to hold the number of leaks in the assembled plant to an absolute minimum. That is, it was most essential to be able to certify that various complex pieces of process equipment would not have to be disassembled for repair in the field, and that the vacuum testing crews in the field could have confidence in the tightness of the component pieces of equipment. The decision to pre-test every piece of process equipment and piping subassembly required the preparation of many varied procedures and specifications, and the education of manpower to use these new techniques in the different places of manufacture. Specifications for each unit were made on the basis of mechanical complexity and the number of such units occurring in each assembled cell of the plant. Table III lists the major components of each cell, the contribution of their volume per one hundred cubic feet of cell volume, and the inleakage contribution and the specified inleakage rate. It should be noted that, if no damage occurred during transport and erection, and if all field welds were tight, the assembled plant would have an inleakage rate considerably below 2 microns pressure rise per hour. Thus a margin was available for depreciation and safety.

Each manufacturer was asked to send several engineers to Kellex for training in the techniques of leak hunting, using the high sensitivity dynamic method and the hood test for the tightness testing. At the conclusion of the training program these engineers were asked to prepare detailed procedures applicable to their particular process unit. These procedures were reviewed by the Kellex Vacuum Engineering Division before they were put into practice. In all cases high $S/V$ ratios were maintained for production testing in keeping with the heavy production schedules. Many jigs and fixtures were designed and built for this purpose; since the undependability of rubber hose and similar makeshift connections was recognized, large sized equipment had flanged connections for testing, while for small equipment the Dresser type coupling proved successful. Best vacuum practice was followed in the construction of testing rigs; packless valves were used and all possible joints were welded, brazed, or hard soldered. In several instances, rigs were equipped with packed valves and sliding jigs; however, these units always proved to have lower sensitivity than those of all metal construction. Figures 5 and 6 show schematically some of the manifolds used for production testing.

### 6. TESTING THE ASSEMBLED PLANT

Delivering sound, vacuum-tight equipment to the plant site was one job. Perhaps a bigger one was to see that this equipment was assembled without injury, and that the assemblies were made tight. This program required the following groups: (1) Vacuum Engineering; (2) Construction Coordination; (3) Vacuum Testing; (4) Acceptance Control and Records.

Vacuum Engineering for the assembled plant was concerned with four major problems:

(1) Definition of the subsections of the plant to be tested as units.
(2) Location of test connections in process system, to assure sufficiently high $S/V$ ratios for efficient testing.
(3) The magnified effect of sorption of the probe gas, when a large number of diffusers were tested simultaneously.
(4) Excessive extraneous leakage through the shaft seals of the centrifugal pumps.

Provision of sufficient test connections on the test cells allowed a careful study of all these problems to be made.

The process design of the plant largely determined the basic sections for testing, since the addition of valves to facilitate testing might interfere with orderly process operation. Thus a cell, building and cell by-pass lines, and the cold traps, were tested as subsections. The limits and state of completion of each subsection were accurately defined.

Each proposed layout of test connections was analyzed using an electrical analogy. The resistance of the piping to gas flow may be considered equivalent to electrical resistance. Consequently, an analysis of resistance network was used to demonstrate the optimum location for the test connections. These analyses were made mathematically or were measured on resistance networks. The results were quantitatively confirmed by later field measurements.

The Construction Coordination Group reviewed the progress of construction and established standards for completion of defined subsections. Briefly, these subsections were defined so that (1) no interference between construction and testing could exist (i.e., all subsections were ended past a limiting valve, so that opening of the valve to permit welding was unnecessary); (2) operations could follow vacuum testing so as to bring the process on stream in a practical orderly manner. Before transfer of buildings or sections to the Vacuum Testing Groups the Construction Coordination Group reviewed the unfinished construction items which might interfere with vacuum testing, and checked the results of preliminary pressure testing. Once a section of the plant was turned over to the Vacuum Testing Group, further construction made necessary by redesign or other reason was subject to the approval of the Construction Coordination Group.

Actual vacuum testing was performed by three-man crews assigned to each leak detector-pump unit combination. These crews were headed by a person of at least high school education and were all given a two-week special training course in the mechanics of the leak testing operation. Because of the complicated nature of the leak detectors and pump units, the test crews were not permitted to make any adjustments or repairs to the equipment. Highly trained personnel was used for the maintenance of all leak detection equipment.

Vacuum Testing Acceptance Control and Record Groups kept a record of all leaks found, the status of testing, and repairs. The results of studies by this group suggested several changes in design, erection procedures, and testing sequence, which considerably reduced both the probability of leaks and the time required for testing. The major function of this group was to review the test data for conformance to specification.

The efforts of the Vacuum Testing Program involving well over 1100 men were climaxed when the plant was brought below a leak rate of 2 microns per hour. This was accomplished in the face of a constantly accelerated construction schedule and achieved within a very narrow margin of time allotted for vacuum testing after the completion of construction.

## 7. APPENDICES

### 7.1. Fundamentals of Vacuum Engineering*

Industrial uses of high vacuum have increased very rapidly in the past few years. Among such applications may be listed the manufacture of radio and x-ray tubes, lamp bulbs, rectifiers, etc., and the reduction of various metal bearing ores.

To satisfy the growing needs of industry in the field, much time and energy have been spent in the design and development of rugged, high capacity pumping equipment, by means of which the pressure in vessels of several hundred cubic feet of volume may be reduced to exceedingly low values with great rapidity.

#### (*a*). *Tightness and Cleanliness*

Tightness and cleanliness are absolutely essential in all vessels used in high vacuum systems. This becomes obvious when one considers that one cubic inch of standard air will occupy a volume of approximately 760,000 cubic inches at one micron pressure, and that one cubic inch of water when vaporized will occupy 945,000,000 cubic inches, at one micron pressure. Unless systems are clean and free from leaks, it is, therefore, extremely difficult to reduce the pressure in them to the high vacuum range.

#### (*b*). *Vacuum Pumping System*

A vacuum pumping system consists of three essential units known as the fore pump, the diffusion pump, and the vapor trap. (See Fig. 3.)

*Diffusion Pump.* The diffusion pump provides a means of obtaining large volumetric pumping capacities in the high vacuum range of pressures. The pumping action originates in the streaming of either oil or mercury vapor from a suitably designed jet. The gases to be pumped diffuse into this vapor stream and are carried along with it away from the system being evacuated. Pumps of this type usually attain their full volumetric capacity at about one micron pressure, and maintain this same value to very low pressures. The capacity of a well-designed pump will be approximately 50 cubic feet per minute per square inch of opening at the jet, and is, therefore, nearly $\frac{1}{3}$ the molecular flow through a hole equal in diameter to the pump housing when absolute vacuum is maintained on one side.

The diffusion pump will not exhaust directly against atmospheric pressure. There is a critical fore-pressure or threshold pressure above which it will not pump at all. This required fore-pressure (about fifty microns) is maintained by a mechanical pump, an ejector, or a combination of the two.

*Fore Pump.* The fore pump is a mechanical pump or ejector. It must have sufficient capacity to handle the gases compressed by the diffusion pump and maintain the fore pressure against which the diffusion pump operates at a value lower than the critical threshold pressure. Mechanical pumps will not in themselves produce pressures lower than about one micron. For lower pressures, the diffusion pump must be added.

*Vapor Trap.* The vapor trap is necessary to prevent the migration of oil vapor from the diffusion pump to the system being evacuated. In most cases, a suitably designed water-cooled baffle will suffice. However, a refrigerated cold trap is required to obtain pressures lower than 0.01 micron, or when it is desirable to prevent small traces of oil from entering the system which is being exhausted. Pressures lower than 0.01 micron may be obtained without the use of cold traps, but only with great difficulty and an excessively long "conditioning" time.

#### (*c*). *Flow of Gases*

The flow of gases in the high vacuum range obeys laws quite different from those familiar to the hydraulic engineer. This fact cannot be stressed too highly since it is overlooked all too frequently by the uninitiated.

Extremely large and short pipes must be used to handle the large capacities of a diffusion pump,

---

* By J. R. Downing, Kellex Corporation. (Present address: Cook Electric Company, Chicago, Illinois.)

TABLE IV. Capacity loss through pipe connections to a diffusion pump (capacity 2600 cfm below 0.1 microns).

| Length of pipe in feet | Percent capacity reduction 4″ I.D. | 8″ I.D. | 12″ I.D. |
|---|---|---|---|
| 2 | 88 | 52 | 27 |
| 6 | 95 | 73 | 45 |
| 20 | 98 | 89 | 71 |

although the resultant mass transfer is very small. For piping systems longer than a few feet, the pipe size used must be of larger diameter than the pump opening. It is customary to calculate pumping system capacities in terms of the capacity of the pump and the flow capacity of the pipe. The expression for over-all pumping capacity is:

$$\frac{1}{S_{\text{system}}}=\frac{1}{S_{\text{pump}}}+\frac{1}{S_{\text{pipe}}}$$

where the various pumping capacities are expressed in cubic feet per minute. If the length of the pipe is more than ten times the diameter, the capacity of the pipe for air is given by:

$$S_{\text{pipe}}=\frac{0.51D^4P}{L}+\frac{14D^3}{L}\,\frac{1+0.62DP}{1+0.76DP},$$

where $D$ is the pipe diameter in inches, $L$ is the length in feet, and $P$ is the average line pressure in microns. A second semi-empirical expression for the capacity of short lengths of pipe at pressures less than 0.1 microns is given by:

$$S_{\text{pipe}}=\frac{14D^3}{L+(D/9)},$$

where $L$, $D$, and $S$ have the same units as in the above formula. It is evident from these expressions that the pipe size may be the limiting factor in the capacity of a pumping system regardless of the capacity of the pump. Table IV gives the reduction in capacity caused by various lengths of pipe connecting a 2600 cfm diffusion pump to the vessel being evacuated.

### (*d*). *Precautions and Techniques*

Precautions and techniques necessary to the attainment and maintenance of high vacuum are so many and varied that the subject will be dealt with here only in the most general way. This is summed up by saying that vacuum systems must be capable of attaining high vacuums without the use of waxes, greases, lacquers, soft solders, and the like.

Sound engineering, cleanliness, ruggedness, and good materials are all of prime importance.

The equivalent in weight and strength of standard 150-pound flanges and fittings is recommended. These may seem excessively heavy but, in view of their reliability and trouble-free operations, they are well worth the additional cost.

### (*e*). *Evacuation Time*

The time required for evacuation may be calculated fairly well for the range of pressures from atmospheric down to a few hundred microns. The expression is:

$$T=\frac{V}{S}\ln\frac{P_1-P_0}{P_2-P_0},$$

where $T$ is time in minutes, $V$ is volume of vessel in cubic feet, $S$ is the pump capacity in cubic feet per minute, $P_1$ is the initial pressure, $P_0$ is the pressure attained by the pump above at blank off, and $P_2$ the final pressure. Since $P_0$ is small compared to $P_1$ and $P_2$, the expression may be simplified to:

$$T=\frac{V}{S}\ln(P_1/P_2).$$

With $P_1$ atmospheric pressure and $P_2$ a few hundred microns, the term involving the logarithm becomes about eight so that the time required for "roughing" a vessel becomes:

$$T=8(V/S).$$

The latter expression holds quite well, provided the system is free from inleakage and the volumetric capacity ($S$) of the pump stays reasonably constant. For most mechanical vacuum pumps the latter condition is true down to pressures of a few hundred microns,

In the range of pressures lower than 100 microns, the above expression is quite useless for three reasons:

1. The capacity of the diffusion pump is by no means independent of pressure.

2. Degassing and vaporization of substances on the inner walls of a vessel contribute so greatly to the volume of gases which must be handled by the pump, that the residual volume of air becomes negligible by comparison.

3. Impedance of connecting pipes becomes an important factor.

If the amount of degassing per unit surface area is known as a function of time and pressure, then the pump-down time may be estimated by a graphical integration method. However, such data are usually not known very accurately so that choice of pump size is dictated largely by experience. The modern trend is to use diffusion pumps of very large capacity so that a high vacuum may be produced quickly and in spite of degassing, small leaks, etc.

### (*f*). *Special Gauges*

Special gauges have been developed for the measurement of pressures lower than one millimeter of mercury. The forces involved are so small that a direct measure of the pressure by observing the deflection of a diaphragm or the equivalent does not prove practical. Some other property of the gas which varies with pressure and which is capable of easy application must be chosen.

The McLeod gauge has long been the standard for the accurate measurement of pressures less than one millimeter of mercury. A sample of the gas in the system is entrapped and compressed by a known ratio. A measure of the resulting pressure then permits calculation of the initial pressure. Compression ratios as high as a million may be used enabling the measurement of pressures as low as 0.001 microns. Because of the high compression ratio, the gauge may only be used for the measurement of pressures due to permanent gases.

The thermocouple and Pirani gauges depend upon the variation of heat conductivity with pressure of the gas. These gauges are, therefore, not absolute and must be calibrated by comparison with the McLeod gauge described above. They are sensitive in the range of pressures between one millimeter of mercury and one micron.

The ionization gauge is capable of measuring pressures from one micron to the lowest attainable which, at present, is about one millionth of a micron. The gauge is essentially a three-element vacuum tube operating with its grid at a positive potential and plate negative with respect to the filament. Electrons from the filament gain energy in their passage to the grid and so will cause ionization of those gas molecules with which they may collide. The resulting positive ions are collected at the plate and constitute the plate current. Since the probability for ionization is proportional to the molecular concentration of the gas, the plate current represents a measure of the pressure.

Ionization gauges, like the thermocouple or Pirani gauges are not absolute, but must be calibrated by comparison with a McLeod gauge.

A number of other gauges have been developed for measuring low pressure, but few if any of them have ever proved practical in industrial applications. There is considerable room for improvement in apparatus for measuring low pressures, since none of the gauges now in use are wholly satisfactory.

## 7.2. Pressure Testing Techniques

Pressure testing for tightness has been applied to vessels in both vacuum and pressure service for many years. Therefore, this section of the report will merely indicate the various methods briefly and discuss their applicability to testing process equipment for high vacuum service. Table V indicates the various methods in general use and indicates the smallest leak that may be detected practically. Extremely careful application of the method will perhaps decrease the quoted figure two- to fivefold.

TABLE V. Pressure testing techniques.

| Test medium | Probe | Response | Leak size at 45 psig, mcfh | Special requirements |
|---|---|---|---|---|
| Air-nitrogen | Flame | Wavering | 5000 | Draft free room. |
| | Sound | Hissing | 5000 | Low noise level. |
| | Soap film | Bubbles | 5 | Each area to be observed under good light for at least five minutes and soap film maintained. |
| | Immersion | Bubbles | 20 | Five to fifteen minutes observation—occluded air on surface must be avoided. |
| Organic halide | Halide torch | Flame color change | 5 | |
| Acidic gas—($CO_2$ or $SO_2$) | Ammonia | Fumes | 5 | |
| Ammonia | $CO_2$; HCl | Fumes | 5 | |
| Liquid under high pressure | | Wet exterior surface | 500 | |

These techniques are used only to locate leaks and do not permit any measure of the amount of leakage to be made. Application of pressure drop measurements are not, in general, suitable to checking the tightness of systems for vacuum service, the permissible drop being beyond the limits of accuracy of the best commercial gauges generally available. This drop for diffusion plant standards would be less than 0.001 lb./sq. in./hr. and corresponds to a temperature change of about 0.005°F. The uncertainty of the temperature correction further invalidates this method for measurement of the vacuum tightness of process vacuum equipment.

After all leaks are thought to have been found, the equipment under test must be evacuated and the rate of pressure rise observed. Consideration of the gas laws reveals that the temperature correction in the absolute pressure range of 1–100 micron is negligible. Vacuum gauges can readily measure the pressure changes required for tightness. However, a serious drawback is the effect of outgassing on the observed leak rate. The subject is more completely discussed in Appendix 7.3.

Other practical drawbacks to the use of pressure testing techniques are (1) the requirement that the system be depressured before repairs are attempted; (2) inability to check hidden leaks (leaks that are not visible on the exterior to the system) such as valve seats and heat exchanger tubes.

### 7.3. Vacuum Testing Techniques Prior to 1943

Many devices used in vacuum practice will not stand internal over-pressures high enough to make pressure testing practical; therefore the use of vacuum testing techniques became necessary in the laboratories and specialty shops dealing with high vacuum equipment. These methods were considered usable only by specialists, and were not considered applicable to commercial process equipment before the inception of the testing program described in this report. The earliest form of vacuum testing was on all glass systems. After evacuation of the system was begun, the discharge of a Tesla coil was played over the glass surface. A leak would show up as a bright spot or "pip." Experienced operators could detect leaks on metal parts of the system by observing the color change of the discharge when the leaks were sprayed with water or acetone. However, considerable uncertainty was attached to this technique, and one often resorted to "painting" the metal system with a vacuum wax, or lacquer, on a wiped layer of solder or lead.

Perhaps the first large scale vacuum testing of metal systems was done by research groups working with cyclotrons and similar equipment used in nuclear research. In general, those methods depend upon the effect on a hot wire gauge of the change of composition of the gas in the system when the probe fluid displaced the air flowing through the leak. Precautions were taken to insure electrical stability and sensitivity of the measuring circuits and mechanical stability of the gauges used. Probe fluids were carefully chosen to insure maximum effect on the gauge. Thus the ideal probe fluid had low viscosity and either high or low thermal conductance compared to air.

When tightness requirements became more severe, the hot wire gauge was replaced with an ion gauge, since the tightness was sufficient to permit the evacuating pumps to reduce the pressure below the operating maximum of this gauge (<0.5 micron). Hydrogen gas or a gaseous hydrocarbon was most generally used for the probe fluid, although good results were reported for other fluids. Ease of ionization of the probe molecules is a determining factor in the choice of the probe fluid, since the ionization produced by electron bombardment is proportional both to the concentration of the gas in the gauge and to the character of the gas.

To obtain the ultimate sensitivity of these methods the dynamic method was discarded and one of the following methods adopted: (1) the "backing space" technique, or (2) the apparent change in rate of pressure build-up when probe gas displaces air flowing through the leak. Both methods achieve the improved sensitivity at the expense of an increase in testing time.

Backing space is the name given to the process of multiplying the effect of the probe gas by permitting the diffusion pump to compress the gas into a dead volume. This is achieved by shutting the valve in the foreline leading to the mechanical pump. The method further requires judgment in the analysis of the signal indicated on the gauge controls. The caliber of the personnel required to assure foolproof operation is quite

TABLE VI. Resumé of vacuum testing methods in use prior to 1943.

| Leak detector | Operating range, microns | Size of leak discernible, mcfh | Commonly used probes | Remarks |
|---|---|---|---|---|
| Tesla coil | 50–1000 | | Tesla coil | Can be used only on glass systems. |
| Discharge tube | 100–1000 | | Acetone; Methanol; $CO_2$; $H_2$ | Residual gases confusing. |
| Hot wire gauge | Less than 100 | 10–1000 | Acetone; Methanol; $H_2$ | Affected by pressure changes and residual vapors. |
| Ionization gauge | Less than 0.5 | 1–100 | Gaseous hydrocarbons; $H_2$; $O_2$ | Affected by pressure changes and residual vapors. |
| Hot wire with backing space | Less than 100 | 50–100 | Gaseous hydrocarbons; $CO_2$; $H_2$ | Time-consuming since method is discontinuous. |
| Ionization gauge with backing space | Less than 0.5 | 1–100 | Gaseous hydrocarbons; $CO_2$; $H_2$ | Time-consuming since method is discontinuous. |
| Hot wire with pressure build-up | Less than 50 | 1–100 | Gaseous hydrocarbons; $H_2$; $CO_2$; masking with vacuum putty | Extensive outgassing required. |
| Ionization gauge with pressure build-up | Less than 0.5 | 1–100 | Gaseous hydrocarbons; $H_2$; $CO_2$; masking with vacuum putty | Extensive outgassing required. |

high—this method having been reported by groups using only post-collegiate personnel.

Analysis of the apparent change in pressure build-up caused by the introduction of probe gas is very time consuming, requiring that a careful schedule for probing be adhered to, and a subsequent analysis of the time-pressure graph in conjunction with this schedule be made. A further drawback of the technique is the entirely nonproductive testing time spent reducing the outgassing of the system under test. This period varies between 8 to 48 hours per unit depending on the volume and area of outgassing surface involved.

Having found all possible leaks, a measure of tightness is taken in the following manner. The equipment under test is valved off from the evacuating pump and the pressure rise of the system observed. Evacuation of the equipment during the leak hunting period in general has reduced the outgassing sufficiently so that the pressure rise is an approximately true measure of the inleakage. Outgassing may be detected on the pressure-time graph of the leak rate by a pronounced change of slope. The true leak is generally taken after this change of slope has been noted. The rate thus measured does not differ appreciably from the rate of the same system after it has been completely outgassed.

Outgassing is a complex and unpredictable effect, as the gas contributing to the effect may be adsorbed, absorbed in solid or liquid solution, evaporated from a liquid, or the product of chemical decomposition.

A resumé of vacuum testing methods prior to 1943 is presented in Table VI.

### 7.4. Relation between the Volume Pumped, Pumping Speed, and the Response to the Probe

Assumed that $L$ standard cubic feet per minute are leaking into a volume of $V$ cubic feet which is being pumped at a rate of $S$ cubic feet per minute at the total pressure existing in $V$.

Then, if at time $t=0$ probing of the leak is initiated, the volume $V$ will:

(1) Gain: $Ldt$ (standard cubic feet of probe gas per differential of time $dt$).

(2) Lose: $pSdt$ (standard cubic feet of probe gas per differential of time $dt$) where $p$ is the partial pressure of probe gas, expressed in atmospheres, in $V$.

The net gain of probe gas in volume $V$ is:

$$d(pV)=(L-pS)dt,$$

or

$$V\frac{dp}{dt}=(L-pS).$$

Integration of this differential equation gives the partial pressure of the probe gas in the volume and accordingly the partial pressure, $p_t$, of the probe gas in the stream passing the leak detector at any time $t$ after probing of the leak has begun:

$$p_t=\frac{L}{S}[1-\exp(-St/V)].$$

If, at time $T$, the probing is discontinued, a similar analysis will show that the partial pressure of the probe gas remaining in the system will be given by the equation:

$$p_T=\frac{L}{S}[1-\exp(-ST/V)]\exp[-S(t-T)/V].$$

In Fig. 2 the response and clean-up to a probe applied to a leak into a volume of forty cubic feet for a period of one second is shown. The response is expressed in terms of percent of the equilibrium response which is given by the equation:

$$p_\infty=L/S.$$

---

# The Magnetic Electron Microscope Objective: Contour Phenomena and the Attainment of High Resolving Power

JAMES HILLIER AND E. G. RAMBERG
*RCA Laboratories, Princeton, New Jersey*

(Received September 3, 1946)

A theoretical and experimental investigation of contour phenomena observed in electron microscope images near focus has been carried out. An explanation for the apparent reversal of the Fresnel fringes at exact focus is given and a method is described for the determination of the degree of asymmetry from the fringe patterns. A procedure for empirically correcting the asymmetries usually present in magnetic electron microscope objectives is outlined and some of the results obtained with a compensated lens are shown. A number of instrumental defects which may prevent the attainment of the ultimate resolving power of an instrument, together with methods for their elimination, are listed.

## I. INTRODUCTION

IN a recent paper[1] it was pointed out that investigations made on the pole-piece system of the objective lens in a magnetic electron microscope operating under optimum instrumental conditions indicated that the practical limit of resolving power of this instrument had been imposed in the past by asymmetries in the lens field. The work also indicated that it may be possible to attain, in practice, the theoretical limit of the existing lenses if those asymmetries could be completely removed.

The present paper is intended as a more complete report on those studies and on others, made subsequently, which have led to a better understanding of the magnetic electron microscope objective, to quantitative methods of testing the quality of a given pole-piece system, and to a method of empirically compensating for the asymmetries in the imaging system of a magnetic electron microscope.

The work has indicated that the larger magnetic electron microscopes in use in this country are inherently capable of resolving powers of the order of 10A when provided with objective lenses of sufficiently high quality. Furthermore, the attainment of such a high level of performance appears to present a parallel to the attainment of the limiting resolving power in the light microscope in that it depends on the skill and experience of the individual microscopist as much as on the availability of particularly good lenses and other equipment. In fact, as will be seen from the following, the methods by which this performance has been achieved in this laboratory

---

[1] J. Hillier, "Further improvement in the resolving power of the electron microscope," J. App. Phys. 17, 307–309 (1946).

are, for the present at least, not dependent on any commercial development and can be repeated in any electron microscope laboratory possessing normal facilities. Because of this, it is the further intention of the authors to present, in as detailed form as possible, procedures and precautions which must be observed in order to achieve the maximum attainable resolving power.

After the instrument used in this laboratory[2] had been "tuned up" to the point where no instrumental defects other than those inherent in the lenses were limiting the performance, at least at the 20A level, it was realized that asymmetries in the field of the objective were the most important cause of the remaining defects in the image and that the appearance of the image contours in the best-focused images was the most sensitive test of these asymmetries. Owing to the refractive effects in a transparent membrane, as discussed below, there is an apparent reversal of the Fresnel fringes as the power of the objective is varied through the focal value. In the best-focused images produced by a lens which is not symmetrical, that is, a lens for which the focal length is not independent of the azimuth of the meridional plane in which any particular paraxial ray is traveling, the reversal in the Fresnel fringes produces a characteristic type of asymmetry which is shown in Fig. 1. This is a high magnification image of an opening in a thin collodion membrane. Here, in one direction (vertical) the contours are those obtained when the power of the objective is less than the focal value, while the contours in the other direction (horizontal) are those obtained when the power of the objective is greater than the focal value. In Figs. 2 and 3, which are images of the same field as in Fig. 1, taken with the same lens, but with objective powers, respectively, much less and much greater than the focal value, the image asymmetries are no longer visible. In practice, the range of the objective current over which the image asymmetries are detectable provides a quantitative measure of the degree of asymmetry existing in the lens. As a further check it is

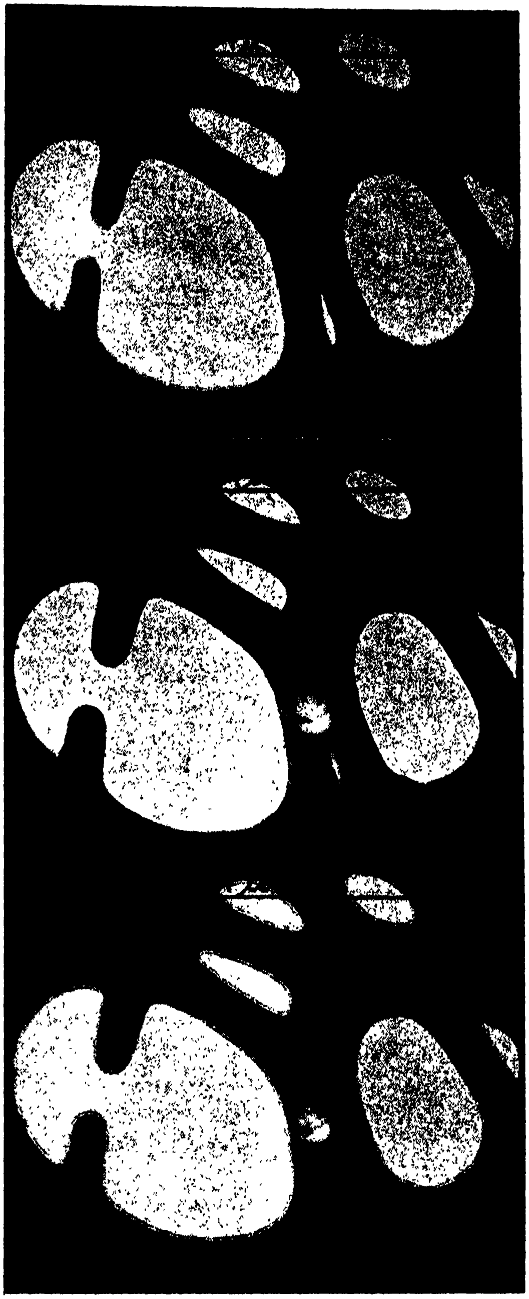


FIG. 1 (upper). Focused image of holes in collodion membrane obtained with asymmetrical objective. Vertical edges show fringes characteristic of under-focused image. See Fig. 2. Horizontal edges show fringes characteristic of overfocused image. See Fig. 3.

FIG. 2 (middle). Underfocused image (focal length too long) obtained with the same objective.

FIG. 3 (lower). Overfocused image (focal length too short) obtained with same objective.

---

[2] An RCA EMU electron microscope using the saturated emission type electron gun and fitted with a telescopic image viewing device (reference 1). With the exception of these differences and those introduced later as a result of the present work, the instrument is a standard commercial model.

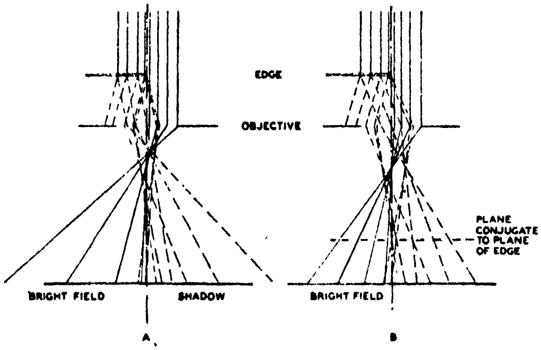


FIG. 4. The imaging of an edge by a parallel beam of electrons: A. Effect of spherical (or chromatic) aberration; B. Effect of defocusing.

possible to estimate the variation of focal length with the azimuth of the meridional plane from a photometric study of the asymmetries in the contours as described below. While important in the academic study of contour phenomena, this latter technique is cumbersome in practice and was not used in the actual compensation of the lenses.

## II. THEORY OF CONTOUR PHENOMENA FOR SMALL ANGLE OF ILLUMINATION

At the edge of an object projecting into space or of a hole in a film the electron intensity distribution in the image ceases to bear a simple relation to the scattering power of the object. Fringe systems are observed whose intensity distribution depends both on the focusing of the objective lens and the condition of illumination. It will here be assumed that a straight edge intersecting the optic axis of the electron microscope at right angles is being observed and that the object is illuminated by a parallel beam of electrons.

Disregarding, to begin with, diffraction phenomena, it is seen that all the electrons passing through the unobstructed part of the object plane fall on the corresponding half-plane of the image, forming there a near-uniform intensity distribution, regardless of the condition of focus and the optical quality of the objective (Fig. 4). The electrons passing through the material portion of the object will be scattered to some extent, so that, except for the position of perfect focus of a perfect lens, a fraction of the electrons leaving the material portion of the object near the edge will fall on the bright half-plane of the image. Thus, for parallel illumination and with diffraction phenomena disregarded, spherical aberration, chromatic aberration, and defocusing all give rise to the same type of contour phenomenon—a bright fringe on the bright side of the edge and a dark fringe, antisymmetric to the bright fringe, on the dark side of the edge (Fig. 5).

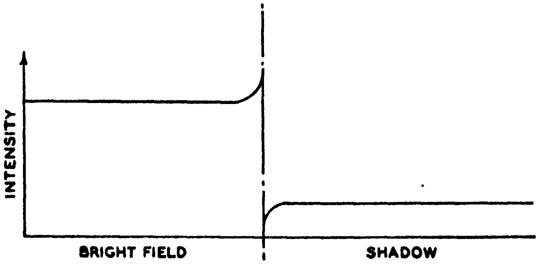


FIG. 5. Intensity distribution at the image of an edge illuminated by a parallel beam, neglecting diffraction effects.

The contour fringes normally observed with small angle of illumination are much more complex than this. This complexity must be ascribed primarily to Fresnel diffraction at the edge. The intensity distribution on a screen placed a distance $l$ behind an opaque object illuminated by a parallel beam of light is shown in Fig. 6. It may be regarded as the result of interference between the unobstructed plane wave, cut off sharply at the geometrical shadow of the edge,

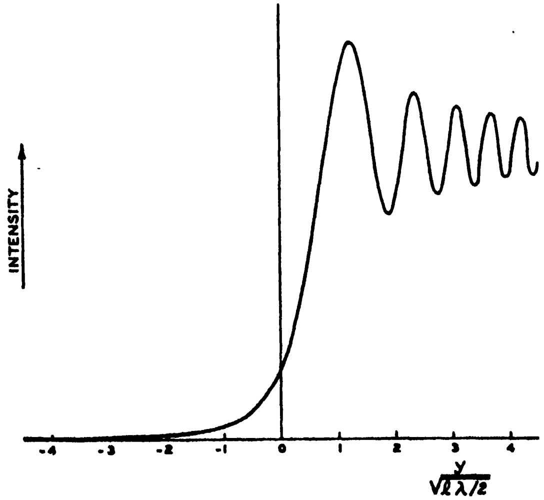


FIG. 6. Intensity distribution behind the edge of an opaque screen illuminated by a plane wave.

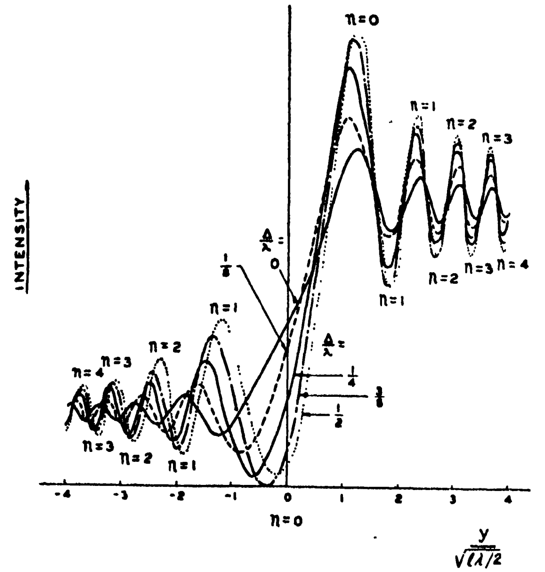


FIG. 7. Intensity distribution behind the edge of a screen transmitting a quarter of the incident radiation with a phase delay $2\pi\Delta/\lambda$ (real pattern).

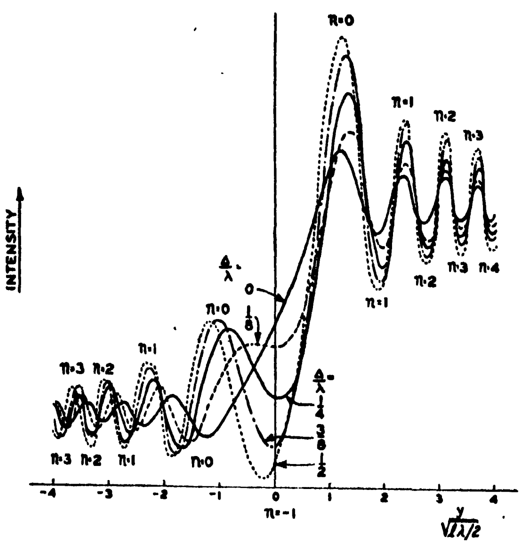


FIG. 8. Virtual intensity distribution ahead of the edge of a screen transmitting a quarter of the incident radiation with a phase delay $2\pi\Delta/\lambda$.

and a cylindrical wave diverging from the edge.[3] The intensity of the edge wave is a maximum in a forward direction and drops off laterally. Its phase is retarded by $\frac{1}{8}$ cycle with respect to the plane wave in the shadow of the edge, advanced by $\frac{3}{8}$ cycle in the bright region; at the shadow of the edge (where this representation ceases to be valid) a phase reversal takes place. In view of these relationships the maxima and minima of the diffraction pattern in the bright part of the field are located at a distance from the shadow of the edge given by

$$y_{\max} = (2l\lambda[n+\tfrac{3}{8}])^{\frac{1}{2}}, \quad y_{\min} = (2l\lambda[n-\tfrac{1}{8}])^{\frac{1}{2}}, \qquad n = 0, 1, 2, 3, \cdots. \tag{1}$$

Here $\lambda$ represents the wave-length of the radiation employed. In the shadow the intensity falls off uniformly.

If a plane ahead of the edge is imaged on a screen by a perfect lens, an identical pattern is observed, though now the fringes result from the interference of the shadow portion of the edge wave with the direct wave. Since the wave advances in phase by a quarter cycle in passing through a line focus[4] the phase of the virtual wave continued backwards from the shadow portion of the edge wave is retarded by $\frac{3}{8}$ cycle with respect to the plane wave. Furthermore, the sign of the path differences for the two interfering waves is reversed in the virtual pattern, so that its maxima and minima have the same location relative to the shadow of the edge (continued backwards) as the maxima and minima of the real pattern. Mathematically, if the amplitude of the wave is given in complex notation, the amplitude in the virtual field ahead of the edge becomes simply the complex conjugate of the amplitude in the real field, an equal distance behind the edge.

If the screen transmits a given fraction $p^2$ of the incident radiation without destroying its coherence or modifying its phase, the expected pattern may be obtained from a superposition of the amplitudes for two plane waves of equal phase differing in initial amplitude by the factor $p$ and incident on complementary opaque screens. Again the expected intensity distribution in a real plane behind the edge and the virtual distribution in a plane at the same distance ahead

---

[3] See, e.g., A. Sommerfeld, "Theorie der Beugung" (especially p. 843) in P. Frank and R. v. Mises, *Die Differentialgleichungen and Integralgleichungen der Mechanik and Physik* (Vieweg, Braunschweig, 1935), part 2, second edition.

[4] For derivation of this "Gouy effect" see P. Debye, "Light near a focal point or line," Ann. d. Physik **30**, 755–776 (1909).

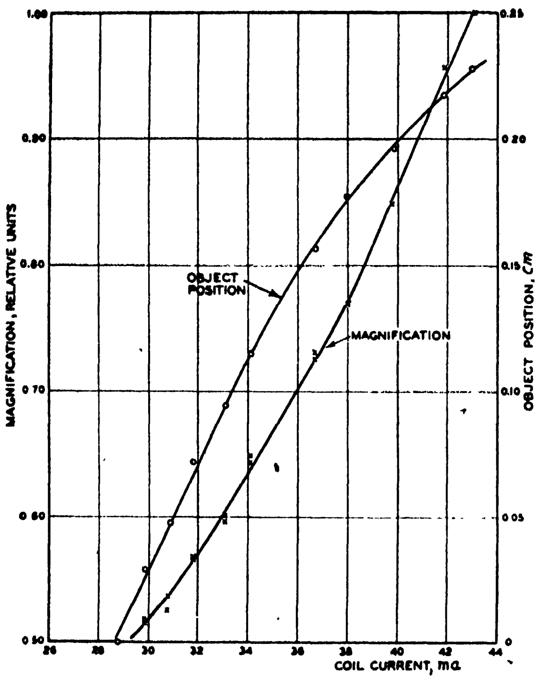


FIG. 9. Measured variation of object position and magnification as function of objective coil current.

of the edge are identical (Figs. 7 and 8, curves for $\Delta/\lambda=0$). It should be noted that in the real pattern the fringe system in the shadow region results from the interference of the wave transmitted through the object with the shadow portion of the edge wave, while in the virtual pattern this fringe system results from interference of the transmitted wave with the part of the edge wave scattered into the bright field.

If the screen with the transmission factor $p^2$ introduces a phase delay in the transmitted wave, the real and virtual diffraction patterns cease to be identical. Let, e.g., the difference in the index of refraction of the screen material and that of surrounding space be $\delta n$ and the thickness of the screen be $t$; then the optical path difference between the wave passing through the screen and that passing through the clear space outside of the screen is $\Delta=t\delta n$. Thus, the phase delay introduced by the screen becomes $2\pi\Delta/\lambda=2\pi t\delta n/\lambda$. The intensity distribution in the diffraction pattern may now be obtained from a superposition of the amplitudes for two plane waves differing in initial amplitude not by $p$, but by the complex quantity $pe^{2\pi i\Delta/\lambda}$, incident on complementary opaque screens. This intensity distribution was calculated, both for the real and for the virtual pattern, for $p=0.5$ and $\Delta/\lambda=\frac{1}{8}$, $\frac{1}{4}$, $\frac{3}{8}$, and $\frac{1}{2}$. The results of the calculations are shown in Figs. 7 and 8. For $\frac{1}{2}<\Delta/\lambda<1$ the real and virtual patterns are identical with the virtual and real patterns, respectively, for $1-\Delta/\lambda$, so that, in effect, the curves in these two figures represent all possible intensity distributions for phase delays which are integer multiples of $\lambda/8$.

The results of the calculations may be summarized as follows:

1. A phase delay enhances the prominence of the diffraction fringes; it becomes greatest for $\Delta/\lambda\rightarrow\frac{1}{2}$.
2. In the range $0<\Delta/\lambda<\frac{1}{2}$ the real pattern contracts to an increasing extent. If the fringes are counted off in the same manner as for the opaque screen (Eq. (1)), the fringes of high order number on the shadow side are displaced, for $\Delta/\lambda=\frac{1}{2}$, by half a fringe toward the edge; on the bright side they are not displaced.
3. For $\frac{1}{8}\leqslant\Delta/\lambda<\frac{1}{2}$ the virtual patterns possess an added minimum and maximum, which may properly be designated as the minimum of order $-1$ and the maximum of order 0 on the shadow side. The high order fringes on the shadow side are displaced a fraction of a fringe away from the edge.

## III. OBSERVED CONTOURS

Contour fringes in out-of-focus electron microscope pictures have been observed and interpreted as Fresnel diffraction fringes by a number

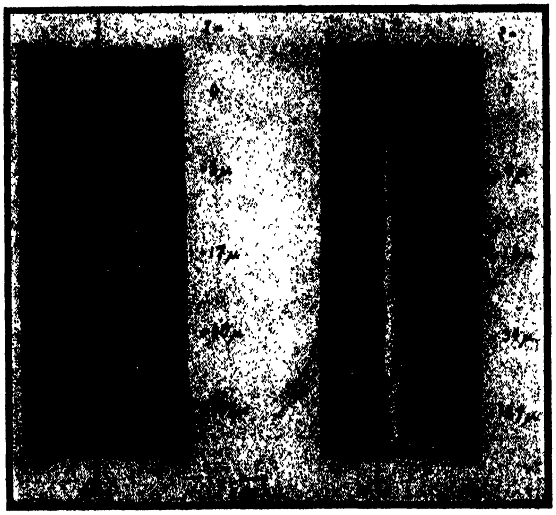


FIG. 10. Electron micrographs of the edge of a collodion film for various degrees of defocusing. The amount of defocusing is given to the right of edge section. Positive and negative values are for over and underfocused images, respectively. The true position of the edge is marked. Compare with Figs. 2 and 3.

of authors.[5-7] Boersch,[7] who made a particularly careful study of the phenomenon, noted that the fringes in the bright field were much more marked in the virtual pattern (objective lens too weak) than in the real pattern (objective lens too strong). He interpreted this as signifying that the shadow portion of the edge wave is much stronger than the edge wave scattered into the bright field. He had reached the same conclusion before from dark field observations.[8] These showed that object edges facing the direction of illumination appeared less bright than the edges remote from the source.

The observations to be presented here were made largely in the hope of shedding more light on the behavior of the fringes near focus. A collodion film with a number of holes of the order of 1 micron in diameter served as object. The illuminating cone was slightly elliptical and had maximum and minimum half-angles of 6 and $4\cdot 10^{-5}$ radian, respectively. These were calculated from the observed image of the effective source, the condenser current, and the geometrical dimensions of the electron microscope which was employed. The variation in position of the object plane of sharp focus with the objective lens current was determined by displacing the object by measured amounts and reading the currents required to bring the image into sharp focus (Fig. 9). It should be mentioned that hysteresis effects in the lens pole pieces cause a variation in the relation between the object displacement and the lens current, so that possibly more accurate estimates of the former may be obtained from the observed diffraction pattern as indicated below. The object displacements as determined by the two methods differed randomly by 10 to 20 percent. The geometrical position of the edge relative to the pattern could be determined roughly, at large deviations from focus, by measuring the magnification of the plate and comparing the known diameter of the hole obtained at focus with the diameters of specific fringes about the hole edges. Results so obtained agreed within experimental error with the edge positions inferred from a comparison of the observed diffraction pattern with that predicted theoretically.

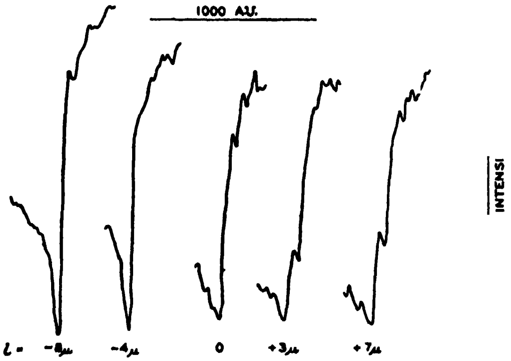


FIG. 11. Microphotometer traces of micrographs of edge taken near focus.

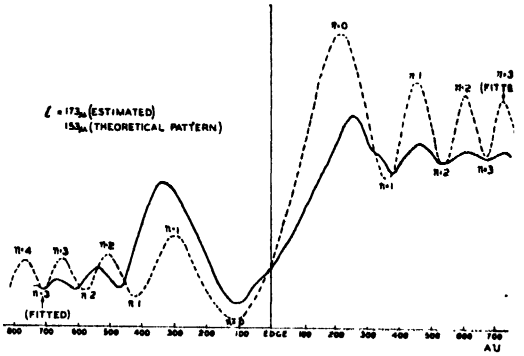


FIG. 12. Comparison of microphotometer trace and Fresnel pattern for edge of screen transmitting a quarter of incident radiation with a quarter cycle phase delay (real pattern).

Figure 10 shows a series of fringe patterns observed for positive (real pattern) and negative (virtual pattern) values of the deviation from focus, as indicated next to the individual micrographs. A vertical line marks the position of the edge. Figure 11 reproduces some microphotometer traces across the pattern for the region close to focus. An inspection of the contour patterns reveals the following:

1. Patterns corresponding to numerically equal deviations from focus in the direction of too great lens currents (real pattern) and too small lens currents (virtual pattern) differ markedly.
2. Proceeding from what, rather arbitrarily, will be

---

[5] E. Ruska, "The origin of fringes about particles in electron microscope images and their change with focus," Kolloid Zeits. 105, 43-52 (1943).

[6] A. Prebus, "The electron microscope," p. 170, in J. Alexander, *Colloid Chemistry* (Reinhold Publishing Corporation, New York, 1944), Vol. 5.

[7] E. Boersch, "Fresnel diffraction in the electron microscope," Physik. Zeits. 44, 202-211 (1943).

[8] H. Boersch, "Edge diffraction of electrons," Physik. Zeits. 44, 32-38 (1943).

referred to as the condition of sharp focus ($l=0$ in Figs. 10 and 11) toward smaller lens currents, an increasingly sharp bright fringe appears on the bright side of the edge, followed by a dark fringe on the dark side. As the lens current is decreased still further, these fringes gradually move apart and added maxima and minima are resolved outside of either.

3. Proceeding from the point of sharp focus to higher lens currents, there appear at the edge a dark fringe and a bright fringe, the bright fringe lying on the shadow side of the dark fringe. These increase rapidly in prominence. Eventually fringe systems become visible outside of these primary fringes, as in the case of too low lens currents. It should be noted that the fringes in the shadow of the edge are much more prominent than in the case of too low lens currents. This is in accord with Boersch's observation that the amplitude of the part of the edge wave scattered into the shadow is greater than that of the part scattered into the bright field. For too high currents (real pattern) the shadow fringes are produced by interference between the transmitted wave and the shadow portion of the edge wave, for too weak lens current (virtual pattern), by interference between the transmitted wave and the portion of the edge wave scattered into the bright field.

## IV. COMPARISON OF OBSERVATIONS AND THEORY

The presence of diffraction fringes in the shadow of the collodion film indicates that the latter has not introduced large *random* phase variations in the electron wave transmitted by it. On the other hand, a comparison of the microphotometer traces of the fringe systems for large deviations from focus with the curves in Figs. 7 and 8 indicates that a satisfactory agreement between the positions of the maxima and minima in the observed and theoretical patterns can only be obtained if the wave passing through the film is assumed to experience a phase delay. The curves for $\Delta/\lambda=\frac{1}{4}$ resemble the observed patterns most closely. Since $\Delta=t\delta n$ and $\delta n=\delta V/(2V)$, where $t$ is the thickness of the film, $\delta V$ the difference in the mean potential within the film and that outside of it, and $V$ is the accelerating voltage of the electron beam, this leads, for $\delta V=10$ volts and $V=50{,}000$ volts, to the perfectly reasonable value for the film thickness, $t=134$A.

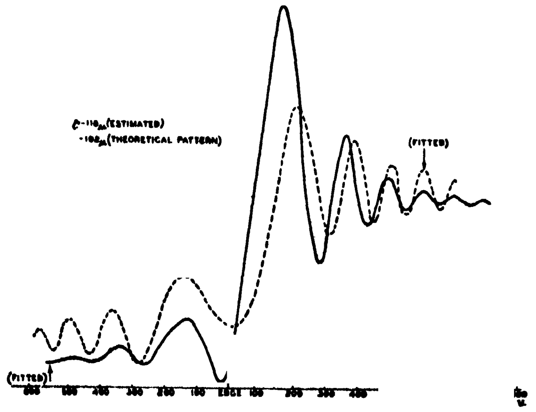


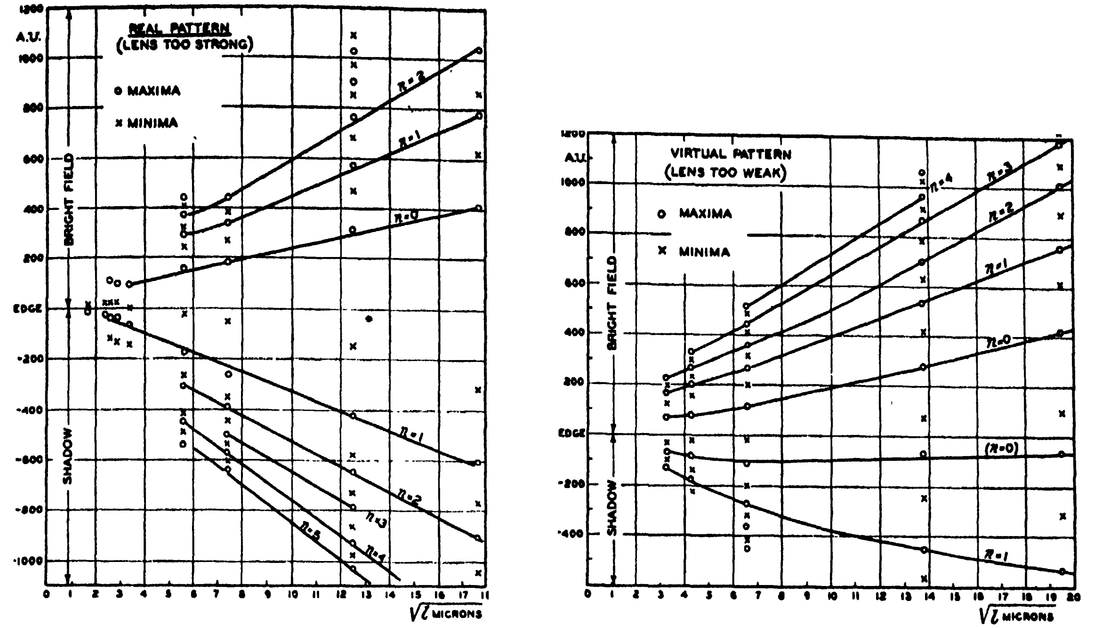
FIG. 13. Comparison of microphotometer trace and Fresnel pattern for edge of screen transmitting a quarter of incident radiation with a quarter cycle phase delay (virtual pattern).

Figure 12 shows a smoothed microphotometer trace of the contour pattern for a deviation, estimated from the lens currents used, equal to 173 microns (lens too strong). In addition, the theoretical curve for $p=0.5$ and $\Delta/\lambda=\frac{1}{4}$ has been plotted as a broken line, adjusting the value of $l$ ($=153$ microns) and the position of the edge so that the maxima $n=3$ on the bright side and the minima $n=3$ on the shadow side of the observed and theoretical curves, respectively, are brought to coincidence. It should be stressed that these curves are comparable only with regard to the positions of the maxima and minima, not to their relative intensities. This is so for the following reasons:

1. The theory here presented does not account for the greater intensity of the shadow portion of the edge wave as compared with the bright-field portion; this appears as a peculiarity of the Fresnel diffraction of electrons by material screens.
2. The finite illuminating aperture employed must lead to a more rapid flattening out of the maxima and minima with increasing order number in the observed curves.
3. The photometer traces shown are simple density plots, no attempt having been made to correct for the nonlinear relation between exposure and density of the photographic material employed.

Figure 13 compares the photometer trace obtained for a virtual pattern (lens too weak; estimated deviation from focus, $-118$ microns) with the theoretical curve for $p=0.5$, $\Delta/\lambda=\frac{1}{4}$. The fitting points were in this case the maximum $n=3$ on the bright side and the minimum $n=2$ in the shadow, the theoretical value of $l$ required to obtain the fit being $-102$ micron.

For both the real and the virtual patterns the correspondence in the positions of the maxima and minima of the observed and theoretical curves, as well as the agreement between the



A. B.

FIG. 14. Variation of distance of maxima and minima from edge with square root of estimated distance between edge and focal plane. A. Real pattern. B. Virtual pattern.

estimated and the theoretical values of the deviations from focus, are close enough to render the interpretation of the observed contour fringes as Fresnel diffraction fringes quite definite.

## V. BEHAVIOR OF THE CONTOURS NEAR FOCUS

If the positions of the maxima and minima relative to the shadow of the edge are plotted against the square root of the separation between the plane of the edge and the plane in which the pattern is observed, these plotted points should fall along straight lines. In practice, deviations are observed, which must in part be ascribed to inaccurate estimates of the distance between the plane of the edge and the plane of focus, as well as to the uncertainty in establishing the position of the geometrical shadow of the edge in the image (Fig. 14). For too large currents the first maximum ($n=1$) and the first minimum ($n=0$) in the shadow can be followed down to the smallest deviations from focus (Fig. 15). Their separation can conveniently be employed as a measure of deviation from focus. Figure 15 shows that these separations are generally slightly larger than the theoretical value for a normal Fresnel pattern, even with a phase delay $\Delta/\lambda = \frac{1}{4}$. For too small currents the maximum $n=0$ in the bright field and the "extra" minimum ($n=-1$) in the shadow are seen farthest. The separation of these two points (Fig. 16) also increases systematically with the deviation from focus, though only relatively slowly.

## VI. IMAGE ASYMMETRIES AND THEIR ORIGIN

### 1. Asymmetries in Electron Micrographs

Contours of the type described above are observed quite generally in electron micrographs; in particular, the character of the fringes nearest the edge is only slightly influenced by the aperture of illumination within the range commonly employed in electron microscopy: The outlying fringes melt into the background as the aperture is increased. Micrographs obtained with an objective lens having some degree of asymmetry will show differences in the fringe patterns surrounding differently oriented edges. These differences are accentuated when the object is focused very carefully (Fig. 1). It will then occur that one edge of a particular object will exhibit the

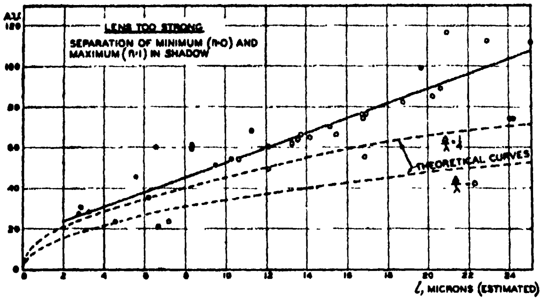


FIG. 15. Separation of first maximum and first minimum in the shadow as function of focal deviation (real pattern—lens too strong).

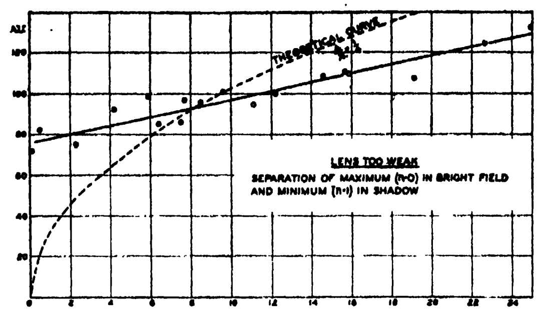


FIG. 16. Separation of first maximum on bright side and first minimum in shadow as function of focal deviation (virtual pattern—lens too weak).

characteristic virtual pattern, corresponding to too low focusing current, while an edge at right angles thereto will appear definitely overfocused. This indicates that the imaging pencils are astigmatic. The degree of astigmatism, i.e., the maximum deviation in focus for two mutually perpendicular directions, can be determined with the aid of the curves in Figs. 15 and 16. The deviation in focus for Fig. 1 may thus be estimated to be 4 microns. Two factors may give rise to asymmetries of the type described: A deviation from axial symmetry of the lens field and a misalignment of the objective.

## 2. Effect of Pole-Piece Ellipticity

Deviations from axial symmetry in the field of a magnetic pole-piece lens may be caused either by imperfections in machining or defects of the magnetic material in the form of inhomogeneity of susceptibility and the presence of blow holes or occlusions. Of these several factors an ellipticity of the pole-piece openings, introduced by imperfect machining, most readily admits of analytical treatment. On the basis of such a treatment it is possible to establish manufacturing tolerances in harmony with a prescribed maximum amount of astigmatism in the image.

The effect on the ray paths of deforming a permanent-magnet ring producing the lens field into an ellipse is derived in Appendix I. The permanent-magnet ring, for which the axial field distribution is shown in Fig. 17, was chosen in preference to the pole-piece lenses usually employed for reasons of ease of calculation. There can be little doubt that the qualitative conclusions—and the quantitative conclusions as far as order of magnitude is concerned—are directly applicable to the types of lenses normally used as electron microscope objectives.

The results of this study, for three objectives of the same focal length, but of varying width of the field distribution (different values of the

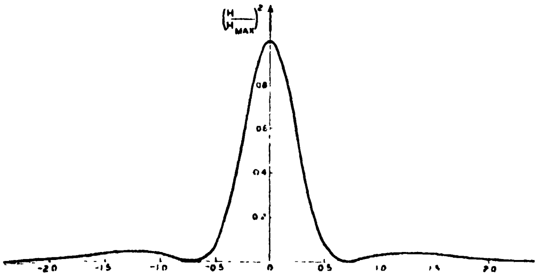


FIG. 17. Axial field distribution of permanent-magnet ring.

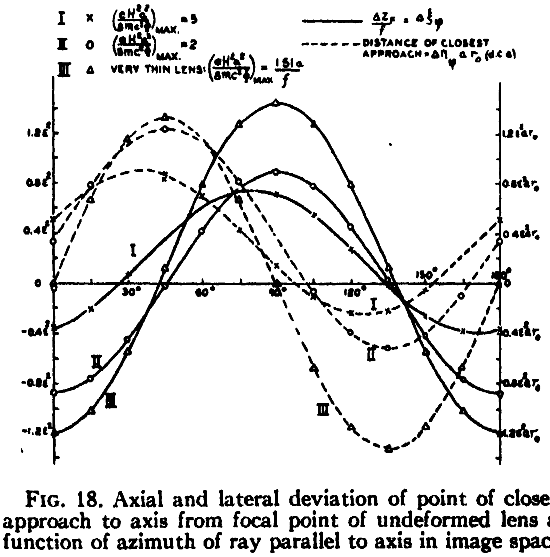


FIG. 18. Axial and lateral deviation of point of closest approach to axis from focal point of undeformed lens as function of azimuth of ray parallel to axis in image space.

TABLE I.

| Len | $\left(\frac{eH^2a^2}{8mc^2\Phi}\right)_{\max}$ | $f$ | $a$ | $s_F$ | $H_{\max}$ | $\chi_a$ | $\left(\frac{\Delta z_F}{f}\right)_{\max}-\left(\frac{\Delta z_F}{f}\right)_{\min}$ | $(d.c.a.)_{\max}$ | $\Delta\varphi$ |
|---|---|---|---|---|---|---|---|---|---|
| I | 5 | 0.3 | 0.435 | 0.162 | 7530 | −47° 33′ | $1.14\epsilon^2$ | $1.57\epsilon^2$ | 10° 7′ |
| II | 2 | 0.3 | 0.265 | 0.235 | 8050 | −40° 50′ | $1.76\epsilon^2$ | $0.79\epsilon^2$ | 38′ |
| III | 0 | 0.3 | 0 | 0.3 | | 0 | $2.64\epsilon^2$ | 0 | 0 |

$H$ = magnetic field
$\Phi$ = accelerating voltage of electrons
$f$ = focal length of lens
$a$ = radius of permanent-magnet ring
$s_F$ = distance of focal point from center of lens
$\Delta z_F$ = longitudinal separation of point of closest approach to axis for deformed lens from focal point of undeformed lens

$d.c.a.$ = distance of closest approach to axis for aperture angle $\alpha = 5\cdot 10^{-3}$ radian
$\chi_a$ = total image rotation
$\Delta\varphi$ = difference in azimuth of parallel ray in image space leading to maximum value of $\Delta z_F$ (weakest lens action) and of the major axis of the ellipse
$\epsilon$ = eccentricity of the ellipse: $\epsilon^2 \cong 2(a-b)/a$, where $a$ is the major and $b$ is the minor axis of the ellipse

radius of the magnetic ring), are summarized in Fig. 18 and Table I. A ray which is parallel to the axis in image space will intersect the axis in object space for only four specific values of the azimuthal angle $\varphi$, measured relative to the meridional plane containing the long axis of the ellipse. For all other values there will exist a distance of closest approach ($d.c.a.$) to the axis, at a point a distance $\Delta z_F$ from the focal point of the undeformed lens (Fig. 19). The ratio of $\Delta z_F$ to the focal length $f$ and the distance of closest approach are plotted as functions of the azimuthal angle in image space, $\varphi$, in Fig. 18. The results shown lead to the following conclusions:

1. An ellipticity of the pole-piece bore results in an astigmatism of the imaging pencils. The distance between the two points of sharpest focus (the focal lines) is proportional to the percentage difference between the two axes of the ellipse. For a focal length of 3 mm a difference in the two axes equal to 0.01 percent ($\epsilon^2 = 2\cdot 10^{-4}$) leads to a separation of the two focal lines of the order of one micron.
2. The ratio of the distance between focal lines and the focal length decreases as the ratio of lens thickness to focal length is increased. This may be ascribed to the fact that, as an imaging ray rotates in the field, the ray is acted upon both by portions of the field which are strengthened and such as are weakened by the deformation. The amount of the rotation is greater for a thick lens than for a thin lens.
3. For weak or short lenses edges in the image which, for proper adjustment of focus, may be made sharp, are parallel to the two axes of the ellipse (edges which appear sharp at the higher lens current are parallel to the minor axis). For strong or thick lenses the two orientations corresponding to the sharpest possible focus are shifted by a small angle with respect to the axes of the ellipse.
4. For weak or short lenses the astigmatic pencil has two focal lines, so that the apparent sharpness of object edges oriented parallel to these, duly refocused, is not reduced by the deformation. For stronger lenses the focal lines become flat ellipses. Under these circumstances—i.e., the presence of "anisotropic astigmatism"—the deformation results in some loss of sharpness throughout the image. The focal ellipses are sufficiently flat, however, that the diffusion of the edges parallel to their axes becomes negligible if the asymmetry in the image is not apparent.

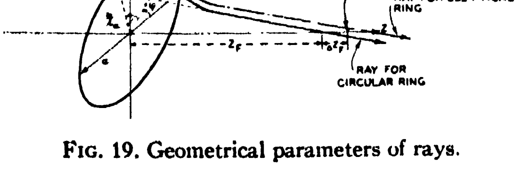


FIG. 19. Geometrical parameters of rays.

### 3. Effect of Misalignment of the Objective

Asymmetries in the image may also be introduced by a horizontal displacement of the objective relative to the projector or a tilt of the objective lens relative to the instrument axis. If $S_4$ is the aberration coefficient of astigmatism[9] for the objective, $M$ its magnification, and $f$ its focal length, the displacement of the two focal lines referred to the object becomes

$$\Delta z_{F\,\max} - \Delta z_{F\,\min} = 2(S_4/M)\cdot f\cdot (d/M)^2 \qquad (2)$$

for a displacement $d$ of the objective relative to the projector and

$$\Delta z_{F\,\max} - \Delta z_{F\,\min} = 2(S_4'/M)\cdot f^3\cdot \delta^2 \qquad (3)$$

for a tilt angle $\delta$ of the objective. The coefficient $S_4$ differs from the coefficient $S_4'$, since the first applies to a position of the effective limiting

---

[9] $S_4 r_0^2 r_a$ is the radius of the circle of confusion owing to astigmatism which forms the image of an object point a distance $r_0$ from the axis, imaged by a pencil whose cross-section radius in the aperture plane is $r_a$. See V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, and A. W. Vance, *Electron Optics and the Electron Microscope* (John Wiley and Sons, Inc., New York, 1945), pp. 551, 562, and 636.

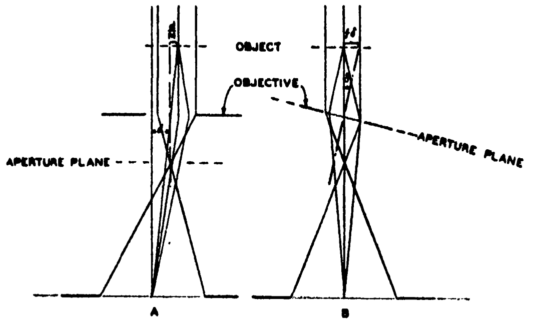


FIG. 20. Passage of imaging pencil through displaced objective (A) and through tilted objective (B).

aperture at the rear focal plane of the objective, the second to one at its nodal point (Fig. 20). To take only one example, the values of $S_4/M$ and $S_4'/M$ for a magnetic objective formed by two unsaturated equidiameter cylindrical pole pieces with a small separating gap are for $\Phi=50$ kv, $f=0.3$ cm, $H_{\max}=6270$ gauss, radius $a=0.575$ cm, which corresponds to a very thick, strong lens,

$$S_4/M=16\ \text{cm}^{-2};\quad S_4'/M=1.5\ \text{cm}^{-2}.$$

For a permissible difference in focus of 1 micron ($10^{-4}$ cm), the tolerance in the displacement $d$ and the tilt $\delta$ become

$$d=\pm 0.3\ \text{cm};\quad \delta=\pm 0.035\ \text{radian}=\pm 2.0°.$$

Both tolerances are seen to be quite large; the normal alignment procedure should render the effects of objective displacement negligible.

## VII. METHOD OF ELIMINATING ASYMMETRIES

When it was realized that the asymmetries in the magnetic field of the electron microscope objective were mainly responsible for the limitations on the resolving power it was assumed at first that these asymmetries were the result of mechanical inaccuracies—elliptical openings and misalignment of the openings of the pole pieces. To overcome these, a lapping technique was devised which insured the required precision of mechanical symmetry and alignment. As will be shown later, mechanical inaccuracies were actually only partially responsible for the asymmetries observed. However, a remarkable improvement[1] was obtained in most cases by the use of the lapping technique and for that reason a brief description is included.

The pole pieces used in these tests are standard objective lens pole pieces for the RCA EMU electron microscope—actually rejects from the commercial production—used *without* an objective aperture. They consist of two soft-iron sections bearing the active pole faces and openings and part of the magnetic circuit of the objective coil unit. The iron sections are threaded into a brass spacer which maintains the proper spacing and alignment through the provision of appropriate pilot surfaces. The cross section of these pole pieces is shown schematically in Fig. 21.

A conical brass lap with a 1°–2° taper was used for lapping the openings. The pole faces were lapped on a plate-glass surface, a large brass jig being used to hold the axis of the lower section accurately perpendicular to the plate. The two openings were lapped simultaneously with the pole pieces assembled in order to obtain accurate axial alignment as well as symmetry. It was necessary, of course, to disassemble the pole pieces in order to gain access to the pole faces for lapping or for cleaning. In order to avoid as much as possible differences in alignment of the two sections occurring as a result of the disassembling, fine scratches were made on the outside of the pole-piece assembly across the boundaries between the pole-piece sections and the brass spacer. Each time the pole pieces were assembled, great care was taken to clean the threads and pilot surfaces and to align the scratches accurately.

The procedure followed was to lap, alternately, lens openings with the lens assembled and faces, individually, with the same grade of grinding compound. A wet paste of 400-mesh carborundum was used for the initial lapping which was continued until all machining marks and other defects disappeared. Lapping was then continued with wet pastes of 600-mesh carborundum and 900-mesh alundum, lapping with the last two grades of grinding compound being continued until the effects of the preceding grade were removed and until there appeared to be a perfect fit between the brass lap and the pole-piece openings for all angular positions of the lap. Considerable care was taken to avoid exerting too much pressure with the lap as this was found to damage the critical edge between the pole-piece openings and the faces. When the grade of grinding compound was changed the pole pieces were very carefully washed to remove all traces of the coarser grade. After the lapping was completed with the 900-mesh alundum and if no imperfections were visible with the 10× eye lens, particularly in the regions near the critical edges, those regions of the pole-piece openings and faces were polished using a wet paste of jeweler's rouge on a round toothpick.

Pole pieces which were treated in this way invariably showed considerable improvement in performance and rather consistently gave limiting resolving powers in the range of 20–30A (Fig. 22). It was soon realized, however, that the asymmetries were seldom completely removed by this technique. Moreover, if considerable asymmetry existed after the first lapping, it was usually found that it persisted upon further lapping. In cases where the asymmetry was found to increase with further lapping, it could generally be shown to arise from defects in the magnetic material of the pole piece, in particular, blow holes or occlusions which occurred in the neighborhood of the critical edge and which would eventually be uncovered. These observations indicated rather conclusively that the asymmetries remaining after lapping were caused by various forms of imperfections in the magnetic properties of the pole-piece material.

Considerable effort was expended in an attempt to obtain ideal magnetic materials. Up to the present none has been found. In fact, the outcome of this work pointed to the empirical correction of the lens asymmetries as being more likely of success and less tedious in execution.

Asymmetries of the magnitude encountered in the uncorrected objective may be considered as a superposition of a cylindrical magnetic lens on an accurately axially symmetrical magnetic lens. The practical method of correction involves, therefore, the production of an additional superimposed cylindrical field which is of the proper strength and at right angles to the existing cylindrical component. There are many possible ways of accomplishing this. The device selected is illustrated in Fig. 23. It consists simply of eight soft-iron rods 0.070 inch in diameter threaded into the wall of the brass spacer (72 threads to the inch), perpendicular to the axis of the lens in the plane of the pole-piece gap and 45° apart. Figure 24 is a photograph of a pole piece with the compensating screws in place.

The procedure of compensation is as follows:

1. A focal sequence (Series 1) is obtained under the conditions to be described below with the rods adjusted so that they are approximately equidistant from the axis of the lens. The orientation and magnitude of the asymmetry is noted on the appropriate exposure (Fig. 25).
2. Any opposite two of the rods are moved as far as possible toward the axis and another focal sequence (Series 2) is obtained under conditions identical to those of Series 1. This operation puts in an extreme degree of asymmetry and permits a correlation to be made between the direction of the asymmetry in the image and the direction of adjustment of the rods in the lens (Fig. 26).
3. The pair of rods which must be moved toward the axis in order to correct for the asymmetry existing in Series 1 is then determined by a comparison of Series 1

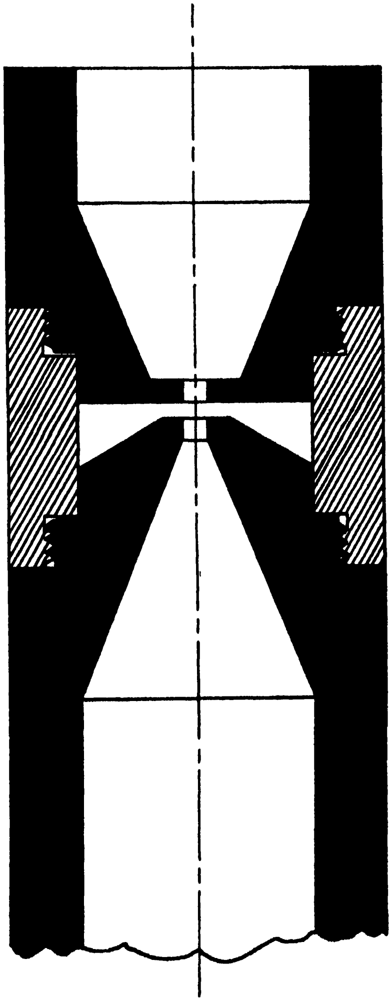


FIG. 21. Cross section of objective pole pieces.

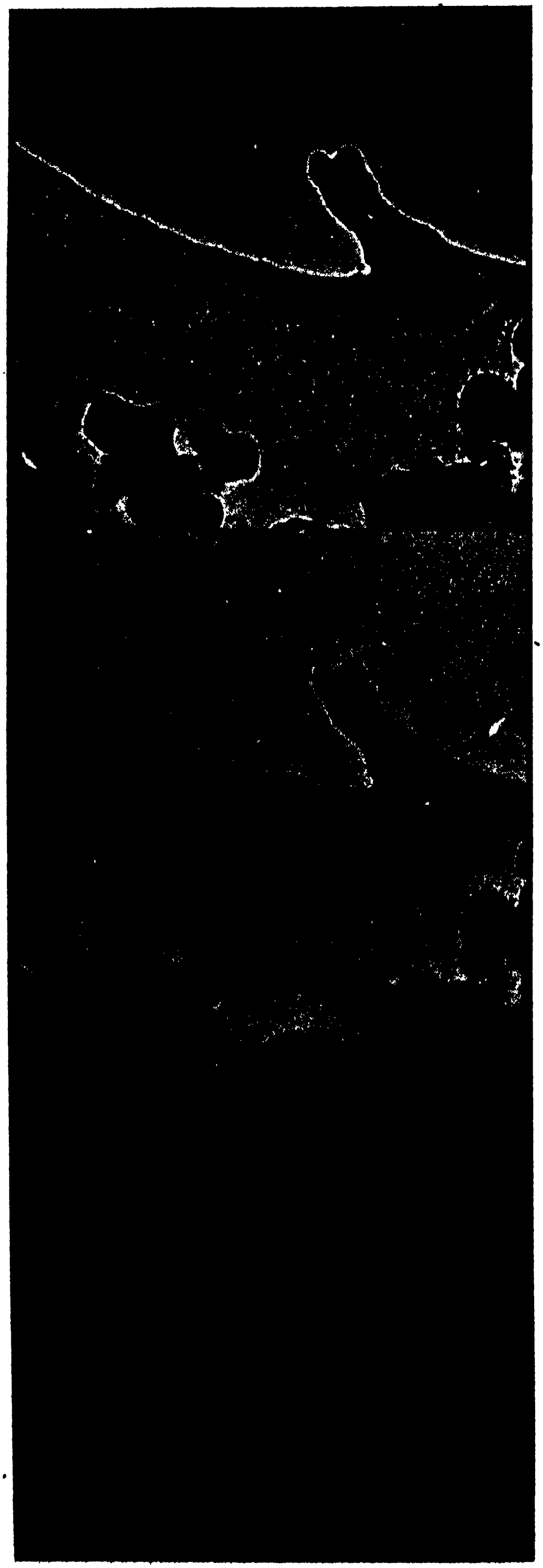


FIG. 22. A focal series obtained with a lapped lens, showing fairly complete elimination of asymmetries. The specimen is a very light evaporated film of gold on a carbon black in collodion mount. This series shows the rapid increase in the diameter of the images of small particles which occurs as the instrument is put out of focus.

and 2. The compensation is of course perpendicular to the pair of rods which would increase the asymmetry observed in Series 1.

4. Starting with the rods equidistant from the axis the two selected for the first correction are moved toward the axis and a third focal sequence (Series 3) is obtained. The direction and degree of asymmetry remaining is determined from the proper image (Fig. 27).

5. If the asymmetry is found to be decreased but has not changed direction the same two rods are moved closer to the axis and another focal sequence (Series 4) is obtained. As shown in Appendix II, the maximum correction attainable by moving any one pair of rods toward the axis is reached when the direction of the asymmetry in the image is rotated through an angle of 45° (Fig. 28).

6. With the rods used in the initial correction set at the point where the direction of the asymmetry has been rotated by an angle in the range of 30° to 45° from its original direction a new direction of correction is determined by comparison of the last series and Series 2. With the rods in the newly selected direction advanced slightly toward the axis another focal series is obtained (Series 5 -- Fig. 29).

7. Correction is continued in the newly selected direction until the direction of asymmetry again changes.

8. This systematic compensation of the asymmetries is continued until no detectable asymmetry exists in the images (Fig. 30).

In actual practice it is found that the asymmetry decreases very rapidly and a third direction of compensation is seldom necessary.

## VIII. INSTRUMENT ADJUSTMENT

In discussing the correction of a lens by the above method no mention has been made of the adjustment of the other components of the electron microscope employed. It is obvious that all the images used in the course of the correcting procedure must be obtained under identical electron-optical conditions. It has been found, furthermore, that the compensation obtained by the above technique is accurate for only *one* adjustment of the instrument. As would be expected, the method described corrects not only the asymmetries inherent in the lens, but also those introduced by stray fields from the other components of the instrument. The various phenomena encountered and the adjustments which must be carried out if the procedure is to give consistent results are listed below:

*1. Specimen.*—A suitable specimen must be used. Small holes in the collodion membrane or spherical carbon-black particles mounted without a supporting membrane have been satisfactory in this laboratory. *Crystalline materials are to be avoided* because of the artificial asymmetries sometimes introduced by different interactions of the electron beam with different faces of the crystals.

*2. Illumination.*—The angular aperture of the illumination should be kept as small as possible in order to obtain the sharpest possible definition in the contour fringes. In the present work the angular aperture of illumination was, in general, below $10^{-4}$ radian, though it might have been as much as an order of magnitude larger without seriously limiting the method. The illumination must arise from a single source.

*3. Lens Position.*—The objective pole-piece assembly must be replaced in exactly the same position in the objective lens coil structure each time it is put back after making an adjustment. This concerns only the angular orientation since the axial position is usually fixed by the structure of the lens.

*4. Specimen Position.*—It is obvious that an empirical method of compensating for lens field asymmetries arising as a result of imperfections in the magnetic properties of the material will be precise for only one condition of magnetization. This means that the specimen position and the accelerating potential of the electrons must be accurately reproduced both during the compensation of the lens and during subsequent high resolution work.

*5. Magnification.*—The preceding condition insures that the objective magnification be constant during its compensation and use. Because of the stray fields produced by the projection lens, and the dependence of the alignment on the power of the projection lens, it was found necessary to keep its magnification constant, preferably at its maximum value. The total instrumental magnification used in the present tests was 16,000× for the objective shown in Fig. 23 and 25,500× for the standard pole pieces used in subsequent tests.

*6. Excitation of the Objective Lens.*—External compensation in a magnetic lens field of asymmetries which are the result of irregularities in the susceptibility of the magnetic material requires, as already mentioned, that the same condition of magnetization be maintained during the compensation and use of the lens. The purpose of this particular section is to point out that the same condition of magnetization cannot be achieved unless the lens current is turned on in such a way as to insure that every region of the material of the pole pieces is at the same point on its hysteresis loop every time the lens is operated, either during the compensation procedure or in use. The procedure which was arbitrarily selected for the present work and which was found to be satisfactory is as follows: The image is focused visually and the objective current turned off. It is left off for about ten seconds to allow the circuits

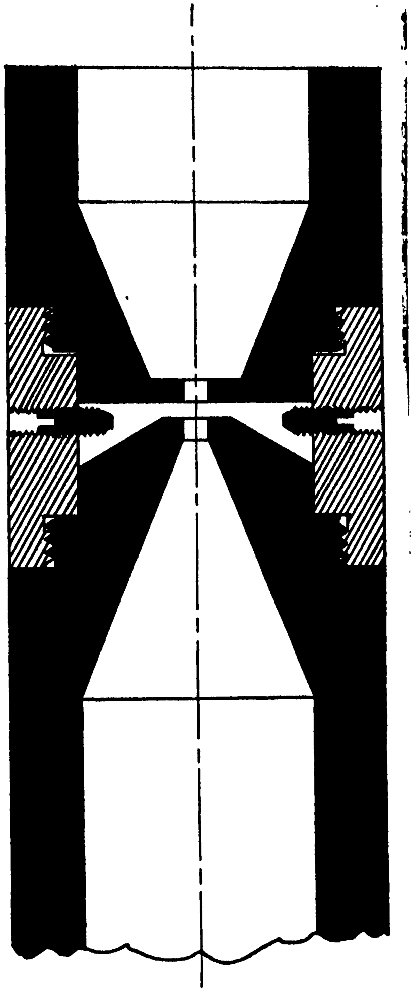


FIG. 23. Cross section of objective pole pieces provided with compensating screws.

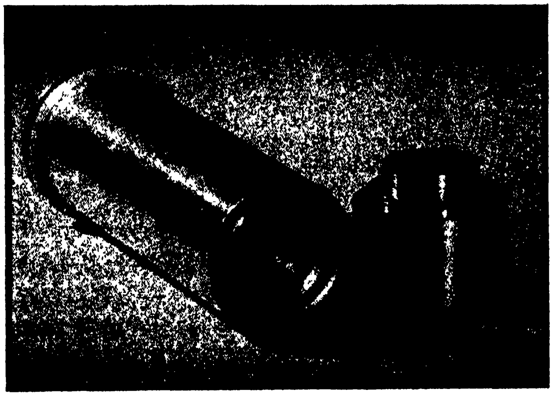


FIG. 24. A photograph of the objective pole piece with the top removed to show the compensating screws.

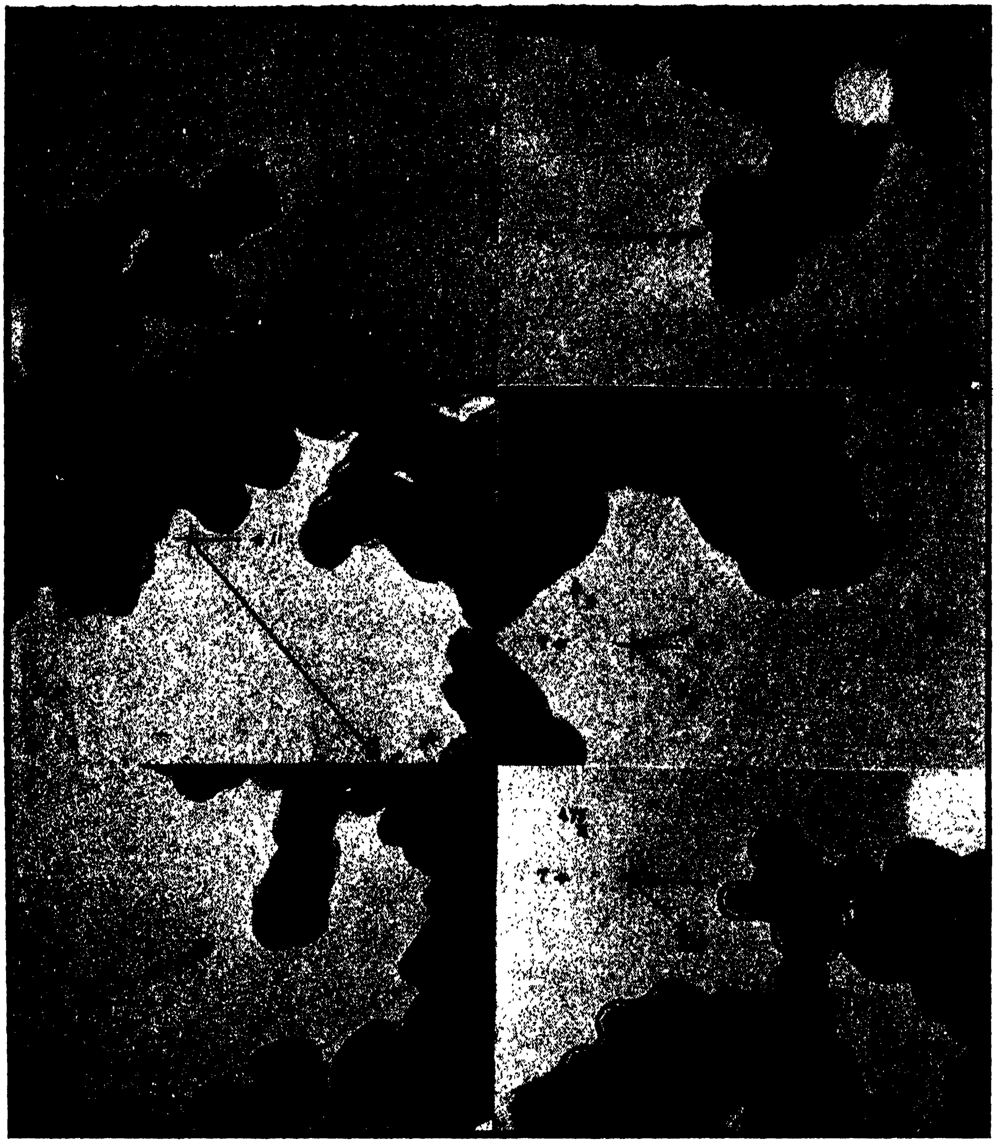


FIG. 25

FIG. 26

FIG. 27

FIG. 28

FIG. 29

FIG. 30

FIG. 25. Focused image obtained with an uncompensated objective.

FIG. 26. Focused image obtained with one pair of correcting screws turned in as far as possible (11 turns). This puts in a large asymmetry which can be correlated with the compensating direction.

FIG. 27. First stage of compensation. The appropriate pair of screws has been selected and has been moved in five turns. The asymmetry is in the same direction as in the initial test (Fig. 25) but of smaller magnitude.

FIG. 28. Illustrates the maximum obtainable compensation with the one pair of screws (7 turns). The magnitude of the asymmetry is greatly decreased and its direction has changed.

FIG. 29. Shows compensation in the selected second direction (4 turns).

FIG. 30. Image obtained with fully compensated objective. The particles in the lower right hand corner are at exact focus. Other particles are either considerably above or below focal plane. Differences in level are emphasized in the compensated objective.

to become stabilized and then turned on again. The image is then observed and refocused. This procedure is repeated until the image comes into exact focus without further adjustment every time the lens current is turned on. The above technique was found to obliterate the previous magnetization history of the pole pieces and to produce consistent results as judged by the observed asymmetries.*

---

* *Note added in proof:* It has been found that the standard objective lens controls were not sufficiently fine for the most critical work. To overcome this a vernier has been

*7. Instrument Alignment.*—As a result of the present investigations the ideas on instrument alignment previously accepted in this laboratory underwent considerable revision. It was found that the quality of the images obtained with a reasonably symmetrical and clean objective lens was insensitive to alignment if considered from the point of view of astigmatism, but was quite sensitive considered from the point of view of chromatic change in magnification and rotation. This is in agreement with theory. The variations in the accelerating potential were measured and found to consist, primarily, of a 3-volt (peak to peak) 60-cycle ripple, which amounts to 0.006 percent of the total voltage and is well within the permitted variations for the attainment of a resolving power of 10A. This tolerance is valid, however, only if the "voltage center" (that is, the point of the image which remains stationary with change in accelerating potential) coincides with the center of the image field. In the present work it was found that, as a result of asymmetries associated with the projection lens, the magnetic and voltage centers were widely separated when the instrument was aligned on the basis of image rotation with objective current change and that the chromatic variation in magnification appeared as a 40A image shift.**

It is obvious from these observations that the alignment should be carried out on the basis of voltage rotation. The alignment procedure actually employed is as follows:

a. Center objective over projector mechanically.
b. Turn the accelerating voltage off and on very rapidly at the lowest projector setting and with the condenser adjusted to maximum power at which an image can still be seen, observing the effect on the image.
c. Translate condenser until center of voltage rotation or expansion is in center of field.
d. Reduce condenser current and translate gun until maximum illumination of field is obtained.
e. Alternately repeat steps c and d until voltage center remains in field for all condenser currents.
f. Raise magnification to maximum value and bring voltage center into field by translating objective.
g. Check e.

When the compensation of the objective was carried out according to the routine outlined above and the instrument adjustments, also listed above, were checked every time the objective lens was inserted for measurement it was found that the compensation was accomplished in accord with theory and that any setting of the lens-correcting screws gave reproducible results. It was found further that under the same condition a fully compensated lens was stable over reasonably long periods. One standard compensated lens maintained its symmetry over a two-week period of constant operation by one operator. In the succeeding two-week period, during which time the instrument was operated by several operators, the same lens became very slightly asymmetrical. However, since accurate records of the results of compensating adjustments in the original correction of the lens had been kept, it was a relatively simple matter to re-attain complete symmetry in the lens.†

## IX. ACCIDENTAL INSTRUMENTAL DEFECTS

Throughout the preceding discussion it has been assumed that the electron microscope being used in these tests performed perfectly within the limitations inherent in its design. Electron microscopists are aware, however, that such is seldom the case in practice and that the attainment of consistently high resolving power is only possible if the operator is continuously on the alert for indications of the presence of any of a large number of possible accidental defects. This situation is accentuated when resolving powers better than 20A are desired. In the following, a number of the more common accidental defects which are particularly important in the high resolving power range will be described and some methods for their detection and elimination given.

### 1. External Magnetic Fields

Commercial electron microscopes are normally carefully shielded for stray alternating magnetic fields, so that the image defects due to such fields are reduced below the 20A level. However, in the attainment of high resolving powers, particularly in the case of unusually bad locations, such stray fields still are an important limiting factor. In this laboratory an image shift of 20A (using an objective lens with a 2.5-mm focal length) was traced to this source.

The conventional electron microscope is most

---

added for which the total variation is 1/10 of that of what was previously the finest control.

** *Note added in proof:* The separation between the magnetic and voltage centers has now been eliminated by the introduction of a soft iron tube immediately above the projection lens pole-piece opening and extending to the top of the coil form. This tube was mounted coaxially with the projection lens system and magnetically isolated from it.

† *Note added in proof:* Since the second correction described above was made, this lens has remained accurately symmetrical over a period of six months during which time the pole pieces have been removed from the instrument and replaced well over 100 times.

sensitive to the effects of alternating magnetic fields in the region of the instrument immediately below the objective lens. Thus, the direction and absolute magnitude (that is, *not* referred to the scale of the specimen plane) of the image shifts resulting from alternating magnetic fields must be constant for all images taken with constant projection lens power and must be independent of the objective lens strength and alignment. These properties are sufficient for the identification of image defects due to alternating magnetic fields.

A single additional shield of small diameter (½ in.) mu-metal tubing placed along the axis of the instrument between the objective and the projection lenses was sufficient to eliminate these defects.

### 2. Specimen Vibration

In the present work some image shifts which were fairly consistent in the 10–20A range and erratic in the 20–40A range were finally traced to vibrations of the specimen holder. Since the resolving power of the instrument for visual observation through the final image telescope[1] at 250,000× is approximately 30A only the larger transitory vibrations could be seen. These were correlated with the boiling of the oil in the diffusion pump and the operation of the shutter.

This vibration, which was extremely troublesome in critical work, was finally eliminated by the simple expedient of lengthening the specimen holder until it just touched the upper face of the objective pole pieces. Unfortunately, under these conditions the stage movement controls had considerable backlash and were extremely difficult to manipulate. Thus, while the technique served for the present work, it is not considered a satisfactory solution of the problem.

### 3. Specimen Stage Drift

In the course of the present work it was found that the specimen stage did not remain motionless after being adjusted but tended to drift by as much as 30A during a 30-second exposure. The effect could be partially eliminated by carefully relieving mechanical strains in the stage mechanism by appropriately adjusting the controls and observing the results at high magnification. The above-mentioned technique of steadying the specimen, with its attendant difficulties, was most successful in eliminating specimen stage drift.

### 4. Specimen Drift (Thermal)

When the angular aperture of the illumination is varied from its maximum value for visual observation to a considerably lower predetermined value for photographic recording the total electron current reaching the specimen remains constant, but its distribution over the specimen changes considerably. This is particularly true in the instrument used for the present work. Since, at maximum angular aperture, the area of irradiation is only 10 microns in diameter, the wire supporting screen is seldom bombarded. On the other hand, at an angular aperture of $5 \cdot 10^{-5}$ radian, at which the micrographs are recorded, the area of irradiation is 330 microns in diameter and, hence, includes several meshes of the wire screen. It is obvious that when the change is made a new condition of thermal equilibrium in the specimen screen must be established and that this may involve a change in position of parts of the specimen screen owing to thermal expansion. It has been found that image shifts due to such thermal movements are very common unless special precautions are taken to avoid them. Woven screens are particularly bad in this respect because individual wires can move independently of the rest. Electrolytically produced copper screens give the best results at the present time, though even these are not always completely satisfactory. The screens must be fairly heavy, which means, of course, that the open area is small (20 percent). Special care must be taken with the clamping of a screen in the specimen holder to insure that it is held tightly around its entire circumference.

### 5. Charging of the Specimen

In accord with observations made with the probe diffraction camera it was found that the area of the specimen irradiated became highly charged *if it did not include a grounded conductor.* While the charging of the specimen does not affect the image quality as far as can be determined by visual observations it does give rise to erratic image shifts, making it impossible to attain high resolution micrographs. These image shifts can be completely eliminated by spreading the beam until one of the screen wires is included in the area of irradiation. Under these conditions the entire area of irradiation is held at the same potential as the screen. The mechanism of this discharging, which appears to be the same as encountered in the use of electron "sprays" in electron diffraction, is not yet understood.

It was also found necessary to take special precautions to clean the specimen screens before use. Large particles of lint or other insulating materials outside the area of irradiation often collect sufficient charge to deflect the beam and produce image shifts.

### 6. Contamination of Specimen Holder

Insulating particles and remnants of previous specimens clinging to the edges of the specimen cap are also a common source of image shifts. The existence of such particles can usually be detected by the occurrence of changes in image position with changes in intensity. In this laboratory the specimen cap and the surface of the hole through the specimen holder are kept polished by frequent cleaning with a very fine grade of steel wool wrapped around a round toothpick. The specimen cap is examined with a 10× eye lens for such contamination before inserting the specimen every time high resolution images are desired. Obviously, care must be taken to avoid leaving any loose threads of the steel wool on the holder.

### 7. Contamination of Objective Lens

Insulating particles clinging to the surfaces of the pole-piece openings introduce both image shifts and image asymmetries and are a continual source of trouble. As in the preceding case the existence of such particles makes itself apparent through a change in image position with change in intensity. The particles can be removed satisfactorily with lint-free lens tissue[10] on a toothpick, a 10× eye lens being used to observe the operation.

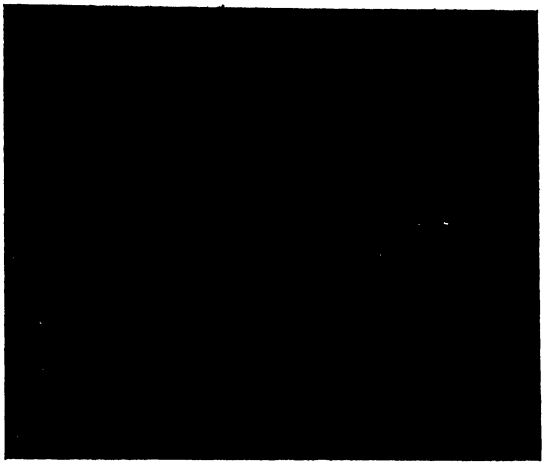


FIG. 31. High magnification image of colloidal gold taken with compensated objective ($f = 3.8$ mm).

### 8. Change of Focus with Change of Illumination

In the early critical work the change of focus which occurs when the intensity and angular aperture of illumination is changed gave considerable difficulty for the narrow range focal series used. The use of the high intensity gun and telescopic viewing device making it possible to focus the instrument and make the photographic exposure without changing the adjustment of the illumination eliminated the difficulty.

### 9. General Remarks

As the reader will have noticed, all of the above defects result in image shifts. If several are present simultaneously and they are observed only in the recorded image it is difficult, if not impossible, to determine their cause. In the present work it has been possible to sort out the various effects and to determine their cause individually through the use of telescopic observation of the image at high magnifications. This permits transitory image motions to be correlated with other incidental phenomena in the operation of the instrument. It is hoped that the presentation of a description of these defects in the above paragraphs will enable other electron microscopists to systematically eliminate them in their own particular instruments, which may not be provided with telescopic observation of the image. It is realized that the above list is not complete and, in particular, does not include

---

[10] Ross-Adams Lens Tissue distributed by Clay-Adams Company, Inc., New York, New York.

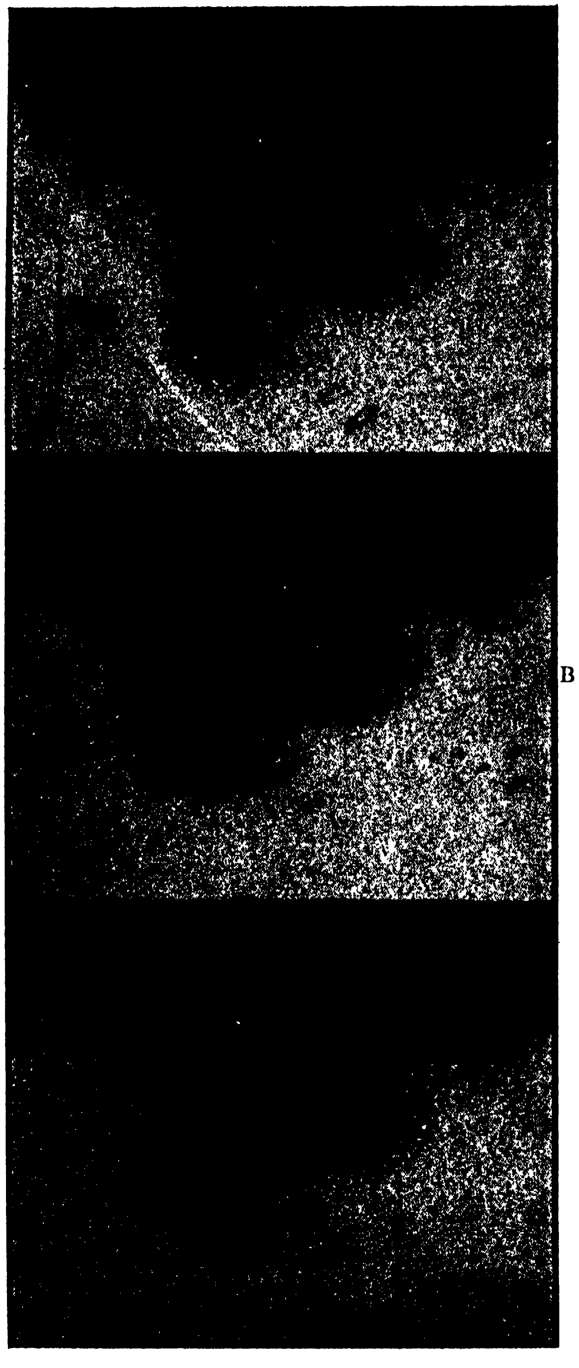


FIG. 32. High magnification focal series of alumina gel taken with compensated objective ($f=2.4$ mm). A. Underfocused image; B. Focused image; C. Overfocused image.

a description of the methods of identifying specific defects in the electrical supplies. The latter can be easily deduced, however, from a study of the normal operation of the instrument.

## X. OBJECTIVE APERTURE

Throughout the development of the electron microscope there have been many differences of opinion with regard to the necessity and value of a physical limiting aperture in the objective. The present work, which was carried out entirely without an aperture, has shown conclusively that such an aperture is not necessary for the demonstration of resolving powers at the 10A level. On the other hand, it has become apparent that lack of contrast at exact focus is at present the major limiting factor in the resolution of extremely fine structures of organic materials. A few attempts, which were made in the course of the present work, to establish the value of a limiting aperture in this particular problem were not conclusive.

## XI. CONCLUSION

It is demonstrated that the Fresnel diffraction fringes which are present in extra-focal images

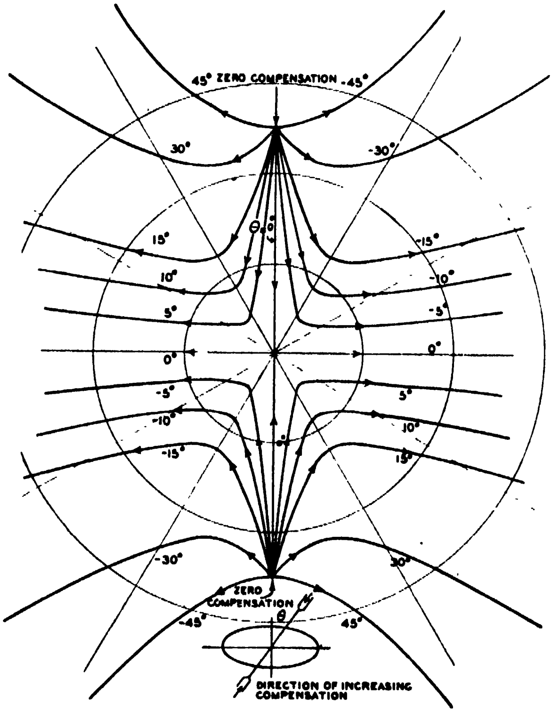


FIG. 33. Variation of orientation and magnitude of residual asymmetry with magnitude of compensating field for various angles between the principal axes of the original field and the compensating field.

obtained with small angular aperture of illumination provide a sensitive criterion of the degree of symmetry possessed by an electron microscope objective. It is shown further that the use of such a criterion makes the correction of asymmetries a relatively straightforward and simple matter. The image quality obtained with a corrected lens (Figs. 31 and 32) has been consistently and obviously higher than that obtained with the best uncorrected lenses.[1] While it has not been possible to make a precise measurement of the best resolving power achieved it is believed to lie in the range of values predicted theoretically for the lenses used.

The observed contour fringes are found to correspond to the Fresnel diffraction pattern for the edge of a partly transparent screen which introduces a phase delay in the coherently transmitted wave. The phase delay, arising from the inner potential of the screen, and the empirically observed greater intensity of the edge wave in the shadow, account for the observed dissimilarity of the contours for underfocusing and overfocusing.

Since the technique of obtaining symmetry in the properties of the image corrects asymmetries owing to external causes as well as those produced by defects in the lenses and since this compensation is not independent of instrument adjustment it appears, at present, that the attainment of high resolving power in a particular instrument requires the individual attention of the electron microscopist responsible for its operation. A number of practical suggestions have been included in the above discussion in the hope that they will assist the interested electron microscopist in the recognition and elimination of some of the more common instrumental difficulties which he may encounter.

---

## APPENDIX I

### Field Variation and Electron Paths for Elliptical Permanent Magnet Ring

If $q$ is the magnetic moment per unit length of the magnetic ring, $a$ (along the $x$ axis) is the major axis, and $b=a(1-\epsilon^2)^{\frac{1}{2}}$ (along the $y$ axis) is the minor axis, the scalar potential variation in the plane containing the major axis is

$$\psi(\xi,0,\zeta)=\frac{q\zeta}{a}\int_{-1}^{1}\frac{(1-\epsilon^2\xi_0^2)^{\frac{1}{2}}d\xi_0}{(1-\xi_0^2)^{\frac{1}{2}}[(\xi_0-\xi)^2+(1-\epsilon^2)(1-\xi_0^2)+\zeta^2]^{\frac{3}{2}}}. \tag{4}$$

Here $\zeta=z/a$, $\xi=x/a$, $x$ and $z$ being the coordinates of the reference point. For the meridional plane containing the minor axis,

$$\psi(0,\eta,\zeta)=\frac{q\zeta}{a}\int_{-1}^{1}\frac{(1-\epsilon^2\xi_0^2)^{\frac{1}{2}}d\xi_0}{(1-\xi_0^2)^{\frac{1}{2}}\{\xi_0^2+([(1-\epsilon^2)(1-\xi_0^2)]^{\frac{1}{2}}-\eta)^2+\zeta^2\}^{\frac{3}{2}}}, \tag{5}$$

$$\eta=y/a. \tag{6}$$

Near the axis a Taylor expansion of the potential yields

$$\psi=\Psi(\zeta)+\frac{1}{2}\left(\frac{\partial^2\psi}{\partial\xi^2}\right)(0,0,\zeta)\xi^2+\frac{1}{2}\left(\frac{\partial^2\psi}{\partial\eta^2}\right)(0,0,\eta)\eta^2+\cdots. \tag{7}$$

This is to be compared with the Taylor expansion for the corresponding axially symmetric system ($\epsilon=0$)

$$\psi_0=\Psi_0(\zeta)-\tfrac{1}{4}\Psi_0''(\zeta)r^2+\cdots, \tag{8}$$

where

$$\Psi_0(\zeta)=\frac{\pi q\zeta}{a(1+\zeta^2)^{\frac{3}{2}}},\quad \Psi_0'(\zeta)=\frac{\pi q}{a}\frac{(1-2\zeta^2)}{(1+\zeta^2)^{5/2}}. \tag{9}$$

For small values of $\epsilon$ the differences of corresponding coefficients in the two expansions are simply proportional to $\epsilon^2$:

$$\frac{\Delta\Psi'}{\Psi_0'}=\frac{1}{\Psi_0'}(\Psi'-\Psi_0')=\frac{\epsilon^2}{4}\frac{4\zeta^4-10\zeta^2+1}{(1+\zeta^2)^{7/2}}, \tag{10}$$

$$\frac{1}{\Psi_{0\max}'}\Delta\left(\frac{\partial^2\psi}{\partial\xi^2}\right)=\frac{1}{\Psi_{0\max}'}\left(\frac{\partial^2\psi}{\partial\xi^2}+\frac{\Psi_0''}{2}\right)=\frac{3\epsilon^2\zeta}{(1+\zeta^2)^{9/2}}\left(\frac{1}{4}-\frac{27}{16}\zeta^2+\frac{\zeta^4}{4}\right), \tag{11}$$

$$\frac{1}{\Psi_{0\max}'}\Delta\left(\frac{\partial^2\psi}{\partial\eta^2}\right)=\frac{1}{\Psi_{0\max}'}\left(\frac{\partial^2\psi}{\partial\eta^2}+\frac{\Psi_0''}{2}\right)=\frac{3\epsilon^2\zeta}{(1+\zeta^2)^{9/2}}\left(\frac{11}{4}-\frac{57}{16}\zeta^2+\frac{\zeta^4}{4}\right), \tag{12}$$

and hence

$$\frac{1}{\Psi_{0\max}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}+\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)=\frac{3}{4}\frac{\epsilon^2\zeta}{(1+\zeta^2)^{9/2}}(12-21\zeta^2+2\zeta^4), \tag{13}$$

$$\frac{1}{\Psi_{0\max}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}-\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)=-\frac{15}{8}\frac{\epsilon^2\zeta}{(1+\zeta^2)^{9/2}}(4-3\zeta^2). \tag{14}$$

The equations of motion of the electrons in the elliptical field are, in the paraxial approximation

$$\frac{\partial^2\xi}{\partial t^2}=\frac{e}{mca}\left(\dot\eta\Psi'-\frac{\partial^2\psi}{\partial\eta^2}\dot\zeta\eta\right);\quad \ddot\eta=\frac{e}{mca}\left(\frac{\partial^2\psi}{\partial\xi^2}\dot\zeta\xi-\dot\xi\Psi'\right);\quad \frac{\partial^2\zeta}{\partial t^2}=0. \tag{15}$$

Here dots indicate differentiation with respect to the time. The equations of motion may be converted into the path equations

$$\eta''=\frac{\ddot\eta\dot\zeta-[(\partial^2\zeta/\partial t^2)\dot\eta]}{\dot\zeta^3}\cong\frac{\ddot\eta}{\dot\zeta^2}=\left(\frac{e}{2mc^2\Phi}\right)^{\frac{1}{2}}\left(\frac{\partial^2\psi}{\partial\xi^2}\xi-\xi'\Psi'\right), \tag{16}$$

$$\xi''=\frac{[(\partial^2\xi/\partial t^2)\dot\zeta]-[(\partial^2\zeta/\partial t^2)\dot\xi]}{\dot\zeta^3}\cong\frac{(\partial^2\xi/\partial t^2)}{\dot\zeta^2}=\left(\frac{e}{2mc^2\Phi}\right)^{\frac{1}{2}}\left(\eta'\Psi'-\frac{\partial^2\psi}{\partial\eta^2}\eta\right), \tag{17}$$

since $\dot\zeta=(2e\Phi/m)^{\frac{1}{2}}/a$ where $\Phi$ is the accelerating potential of the electrons. Primes indicate differentiation with respect to $\zeta$. Equations (16) and (17) may be combined into a single equation:

$$(\xi+i\eta)''+i\left(\frac{e}{2mc^2\Phi}\right)^{\frac{1}{2}}\left[\Psi'(\xi+i\eta)'-\left(\frac{\partial^2\psi}{\partial\xi^2}\xi+i\frac{\partial^2\psi}{\partial\eta^2}\eta\right)\right]=0 \tag{18}$$

or

$$(\xi+i\eta)''+i\left(\frac{e}{2mc^2\Phi}\right)^{\frac{1}{2}}\left[\Psi_0'(\xi+i\eta)'+\frac{\Psi_0''}{2}(\xi+i\eta)\right]$$
$$=-i\left(\frac{e}{2mc^2\Phi}\right)^{\frac{1}{2}}\left\{\Delta\Psi'(\xi+i\eta)'-\left[\xi\Delta\frac{\partial^2\psi}{\partial\xi^2}+i\eta\Delta\frac{\partial^2\psi}{\partial\eta^2}\right]\right\}. \tag{19}$$

Introduce now the new independent variable

$$u=(\xi+i\eta)e^{-i\chi};\quad \chi=-\int_{-\infty}^{\zeta}\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\Psi_0'd\zeta. \tag{20}$$

The differential Eq. (19) now becomes:

$$u''+\frac{e\Psi_0'^2}{8mc^2\Phi}u=-i\left(\frac{e}{2mc^2\Phi}\right)^{\frac{1}{2}}\left\{\Delta\Psi'\left(u'-i\left[\frac{e}{8mc^2\Phi}\right]^{\frac{1}{2}}\Psi_0'u\right)-e^{-i\chi}\left[\xi\Delta\frac{\partial^2\psi}{\partial\xi^2}+i\eta\Delta\frac{\partial^2\psi}{\partial\eta^2}\right]\right\}=R(\zeta). \tag{21}$$

The corresponding homogeneous equation,

$$u_0''+\frac{e\Psi_0'^2}{8mc^2\Phi}u_0=0, \tag{22}$$

has the particular solution $u_\alpha$ with the initial conditions $u_\alpha=0$, $u_\alpha'=-a/f$ for $z=z_F$; for $z=-\infty$, $u_\alpha=1$.

Consider now a ray incident parallel to the axis, unit distance from it, and in a meridional plane with the azimuth $\varphi$. In the zero approximation its coordinates are then given by

$$u=u_\alpha e^{i\varphi};\quad \xi=u_\alpha\cos(\varphi+\chi);\quad \eta=u_\alpha\sin(\varphi+\chi). \tag{23}$$

The deviation of the ray passing through the deformed lens in the plane of focus ($z=z_F$) is then[11]

$$\Delta u=\frac{f}{a}\int_{-\infty}^{\zeta_F}Ru_\alpha d\zeta. \tag{24}$$

Here, to obtain a first approximation, the coordinates in Eq. (23) are to be substituted for $u$, $\xi$, and $\eta$ in the expression $R(\zeta)$ from Eq. (21). It is desirable to refer this displacement to the plane in which the corresponding ray for the symmetrical lens traverses focus. The coordinates are then given by

$$\Delta\xi_\varphi+i\Delta\eta_\varphi=\Delta ue^{-i(\varphi-\chi_a)};\quad \chi_a=-\int_{-\infty}^{\zeta_F}\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\Psi'd\zeta. \tag{25}$$

$\chi_a$ is the total rotation of the image. $\Delta\xi_\varphi$ now represents the deviation parallel to the meridional plane, $\Delta\eta_\varphi$ that at right angles thereto. Substituting $R$ from Eq. (21) in Eq. (24),

$$\Delta\xi_\varphi+i\Delta\eta_\varphi=-\frac{2if}{a}e^{i\chi_a}\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\int_{-\infty}^{\zeta_F}\left[\Delta\Psi'\left(u_\alpha'u_\alpha-i\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\Psi'u_\alpha^2\right)\right.$$

$$\left.-e^{-i(\varphi+\chi)}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}\cos(\varphi+\chi)+\Delta\frac{\partial^2\psi}{\partial\eta^2}\sin(\varphi+\chi)\right)u_\alpha^2\right]d\zeta. \tag{26}$$

$$\Delta\xi_\varphi=\frac{f}{a}\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\Psi_{\max}'\left\{2\sin\chi_a\int_{-\infty}^{\zeta_F}\frac{\Delta\Psi'}{\Psi_{\max}'}u_\alpha u_\alpha'd\zeta\right.$$

$$2\cos\chi_a\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\Psi_{\max}'\int_{-\infty}^{\zeta_F}\frac{\Psi'\Delta\Psi'}{\Psi_{\max}'^2}u_\alpha^2d\zeta+\sin\chi_a\int_{-\infty}^{\zeta_F}\frac{u_\alpha^2}{\Psi_{\max}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}+\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)d\zeta$$

$$\cos2\varphi\int_{-\infty}^{\zeta_F}\frac{u_\alpha^2}{\Psi_{\max}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}-\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)\sin(2\chi-\chi_a)d\zeta$$

$$\left.-\sin2\varphi\int_{-\infty}^{\zeta_F}\frac{u_\alpha^2}{\Psi_{\max}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}-\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)\cos(2\chi-\chi_a)d\zeta\right\}. \tag{27}$$

---

[11] See e.g., E. P. Adams, *Smithsonian Mathematical Formulae* (Smithsonian Institution, Washington, D. C., 1922), Formula 8.410.

$$\Delta\eta_\varphi=\frac{f}{a}\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\Psi_{\max}'\Bigg\{-2\cos\chi_a\int_{-\infty}^{\zeta_F}\frac{\Delta\Psi'}{\Psi_{\max}'}u_\alpha u_\alpha'd\zeta$$

$$-2\sin\chi_a\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\Psi_{\max}'\int_{-\infty}^{\zeta_F}\frac{\Psi'\Delta\Psi'}{\Psi_{\max}'^2}u_\alpha^2d\zeta-\cos\chi_a\int_{-\infty}^{\zeta_F}\frac{u_\alpha^2}{\Psi_{\max}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}+\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)d\zeta$$

$$-\cos2\varphi\int_{-\infty}^{\zeta_F}\frac{u_\alpha^2}{\Psi_{\max}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}-\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)\cos(2\chi-\chi_a)d\zeta$$

$$+\sin2\varphi\int_{-\infty}^{\zeta_F}\frac{u_\alpha^2}{\Psi_{\max}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}+\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)\sin(2\chi-\chi_a)d\zeta. \quad (28)$$

The distance between the focal plane and the plane in which the ray approaches closest to the axis is given by

$$\Delta z_F=-a\Delta\xi_\varphi/u_\alpha'=f\Delta\xi_\varphi; \quad (29)$$

$ar_0\Delta\eta_\varphi$, $r_0$ signifying the distance of the ray from the axis in image space, on the other hand, gives the deviation from the axis at the point of closest approach. Equations (27) and (28) were evaluated by numerically solving Eq. (22) for $u_\alpha$ and performing the indicated quadratures. The results are shown in Table I and Fig. 18.

## APPENDIX II

### Compensation of Objective Asymmetries

Equation (7), representing the axial potential for any pole-piece assembly with two meridional planes of symmetry passing through the $\xi$ and $\eta$ axes, may be written

$$\psi=\Psi_0+\left(\frac{\partial^2\psi}{\partial\xi^2}\right)\xi^2+\left(\frac{\partial^2\psi}{\partial\eta^2}\right)\eta^2+\cdot$$

$$=\Psi_0+\frac{1}{2}\left[\frac{\partial^2\psi}{\partial\xi^2}+\frac{\partial^2\psi}{\partial\eta^2}\right](\xi^2+\eta^2)+\frac{1}{2}\left[\frac{\partial^2\psi}{\partial\xi^2}-\frac{\partial^2\psi}{\partial\eta^2}\right](\xi^2-\eta^2)+\cdots. \quad (30)$$

The last term in this expression is responsible for the astigmatism of imaging pencils refracted by the system.

Assume now that a second asymmetry is introduced, e.g., by a pair of iron screws placed symmetrically with respect to the axis. Let the principal axes of this "compensating" system ($\xi_1$, $\eta_1$) form an angle $\Theta$ with the $\xi$, $\eta$ axes:

$$\xi_1=\xi\cos\Theta-\eta\sin\Theta,\quad \eta_1=\xi\sin\Theta+\eta\cos\Theta,\quad \xi_1^2-\eta_1^2=\xi^2\cos2\Theta-2\xi\eta\sin2\Theta-\eta^2\cos2\Theta. \quad (31)$$

The total asymmetry term resulting from the superposition of the two fields is now

$$\Delta\Psi=A[\xi^2(1-k\cos2\Theta)-2k\xi\eta\sin2\Theta-\eta^2(1-k\cos2\Theta)], \quad (32)$$

where

$$A=\frac{1}{2}\left(\frac{\partial^2\psi}{\partial\xi^2}-\frac{\partial^2\psi}{\partial\eta^2}\right);\quad k=\left(\frac{\partial^2\psi_1}{\partial\xi_1^2}-\frac{\partial^2\psi_1}{\partial\eta_1^2}\right)\Big/\left(\frac{\partial^2\psi}{\partial\xi^2}-\frac{\partial^2\psi}{\partial\eta^2}\right).$$

The subscript 1 refers to the compensating system. If the factor $k$, normally a function of $\zeta$, is a constant, perfect compensation is possible. This occurs, of course, for $\Theta=0$ and $k=1$ and corresponds to a compensating field whose asymmetry terms are, except for sign, identical throughout with those of the original field.

If the factor $k$ is not a constant, the adjustment of $\Theta$ and the magnitude of $k$ (for a short magnetic or for an electrostatic lens) will still bring about correction, though now the imagining pencils can be made stigmatic only in the first order. The image formation in such systems is analogous to the formation of an image focused sharply in two perpendicular directions by two cylindrical lenses with their axes at right angles to each other. For a long magnetic lens compensation will generally be incomplete, since the adjustment of the orientation and the magnitude of an arbitrary compensating field will not suffice to balance both the isotropic and the anisotropic component of the astigmatism resulting from the original field. Below the case of the short lens, for which a single, effective, value of $k$ can be defined, will be treated.

Since, in practice, the orientation of the asymmetry in the original pole pieces is initially either unknown or known only approximately, the principal axes of the compensating field first introduced will, in general, form an angle $\Theta$ different from zero with the principal axes of the original field. It is hence of interest to know how the asymmetry in the image will change with the value of $k$ for different values of $\Theta$.

For any value of $k$ and $\Theta$ the principal axes ($\xi_2$, $\eta_2$) of the resultant field (determining the directions of maximum focal deviation or most pronounced defocusing fringes in the image) will be rotated relative to the principal axes $\xi$, $\eta$ of the original field by an angle $\vartheta$:

$$\xi=\xi_2\cos\vartheta-\eta_2\sin\vartheta,\quad \eta=\xi_2\sin\vartheta+\eta_2\cos\vartheta. \tag{33}$$

Substituting this relation in Eq. (32) and adjusting $\vartheta$ so that the coefficient of $\xi_2\eta_2$ vanishes leads to

$$\Delta\Psi=A(1-2k\cos 2\Theta+k^2)^{\frac{1}{2}}(\xi_2{}^2-\eta_2{}^2), \tag{34}$$

$$\tan 2\vartheta=-k\sin 2\Theta/(1-k\cos 2\Theta). \tag{35}$$

Here the coefficient in Eq. (34) measures the asymmetry of the resultant field (or the value of the eccentricity factor $\epsilon^2$), $\vartheta$ the rotation of the direction of maximum lens action (or of the orientation of the most pronounced fringes) as $k$ (or the compensating field) is increased from zero to the given value. These quantities are plotted, for a series of values of $\Theta$, the angular deviation between the principal axes of the two superposed fields, in the polar diagram shown in Fig. 33. Each curve corresponds to a range in $k$ from 0 to 2; arrowheads indicate the direction of increasing $k$.

Correction takes place only if $\Theta$ is less than 45° in absolute value. If this condition is fulfilled, increasing the compensating field from zero results in an initial reduction in the degree of asymmetry and a subsequent increase. The value of $\vartheta$ passes simultaneously from 0 to $\mp(\pi/2-|\Theta|)$ if the sign of $\Theta$ is $\pm$. Only for $\Theta=0$ is full correction attained; for $|\Theta|=5°$ the asymmetry can be reduced to less than a fifth, for $|\Theta|=10°$, to a third, and for $|\Theta|=15°$, to a half.

# An Investigation of a Method for the Analysis of Smokes According to Particle Size[1]

WILLIAM N. LIPSCOMB,[2] T. R. RUBIN,[3] AND J. H. STURDIVANT
*California Institute of Technology, Pasadena, California*[4]

(Received August 2, 1946)

An electrical precipitation method for the analysis of smokes according to particle size has been investigated. A thin stream of smoke particles moving under laminar conditions in a wind tunnel of rectangular cross section is electrically charged in a small region. The charged particles are then precipitated by the uniform field existing between the charged upper plate and the grounded lower plate of the wind tunnel as a long track along the lower plate. The size-frequency distributions of the particles precipitated at various points on the lower plate were determined from measurements on enlargements of electron-microscope photographs of samples of the smoke precipitated at various points. The agreement between theory and experiment is fair, but the range of sizes of the particles precipitated at a given point is rather large.

## INTRODUCTION

THE use of a precipitation method for the estimation of sizes of smoke particles was first described by Rohmann.[5] His apparatus consisted of a wind tunnel of rectangular cross section with an electric field between the negatively charged upper plate and the grounded lower plate. The smoke was introduced into the wind tunnel through a small tube and was charged in a small region by a corona discharge from a point. The charged particles were precipitated by the uniform field as a track about 5 cm in length. The size of the particles was measured with an optical microscope and compared with the particle radius calculated from the theory of precipitation of charged spherical particles in a uniform field. Rohmann did not report the size-frequency distributions of the particles precipitated at various points, but only the average size in the densest region of the precipitate. No statement was made regarding the regularity of shape of the metallic oxide particles in the smokes which he used for his investigation. Rohmann mentions two difficulties encountered in his experiments. The first was the difficulty of obtaining laminar gas flow in the wind tunnel; even in the absence of a charging corona he was forced to work with gas velocities of nearly 40 cm/sec. in order to obtain a stable thread of smoke. The second difficulty was the disturbance of the thread of smoke by the electrical wind which accompanies corona discharge; he was therefore constrained to use comparatively small charging currents.

The present investigation consists of a more extensive study of the precipitation method for the analysis of smokes according to particle size. A modified form of the apparatus described by Rohmann was used. The rectangular wind tunnel was redesigned in order to give essentially laminar flow at gas velocities of approximately 10 cm/sec. Since the point corona was found to be unsatisfactory, other methods of obtaining charging currents were investigated. The electron microscope was used to record the sizes of particles precipitated at various points in the apparatus. Had the separation of the smoke particles proved more satisfactory a much more rapid method was to be used (based on calibration with the electron microscope results) for the determination of the size-frequency distribution function; this proposed method consisted of measurement of the current carried by the charged smoke particles to small metal collectors mounted in the plane of the lower plate and insulated from it. Because the results were not completely satisfactory, however, only the studies of various charging devices, representative size-frequency distribution curves, and com-

---

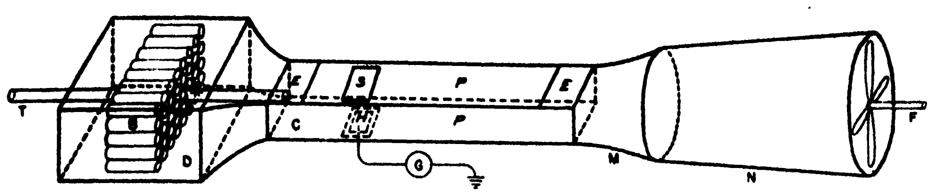
[1] This work was done in whole under the Contract No. OEMsr-103, Supplement 1, between the California Institute of Technology and the Office of Scientific Research and Development, which assumes no responsibility for the accuracy of the statements contained herein.

[2] Present address: School of Chemistry, University of Minnesota, Minneapolis 14, Minnesota.

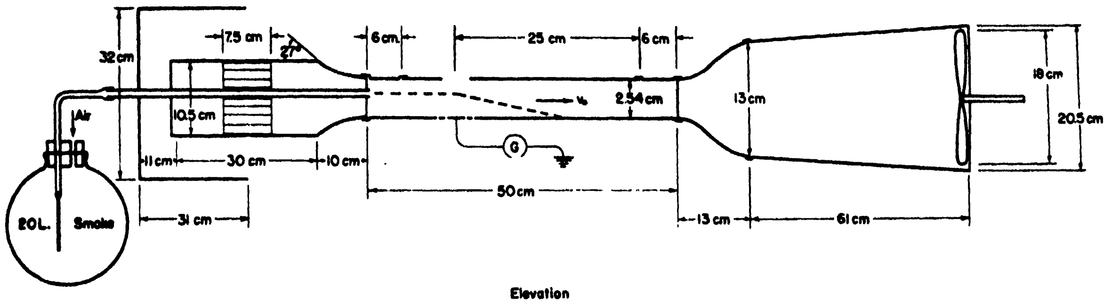
[3] Present address: Department of Chemistry, Ohio State University, Columbus, Ohio.

[4] Contribution No. 1054.

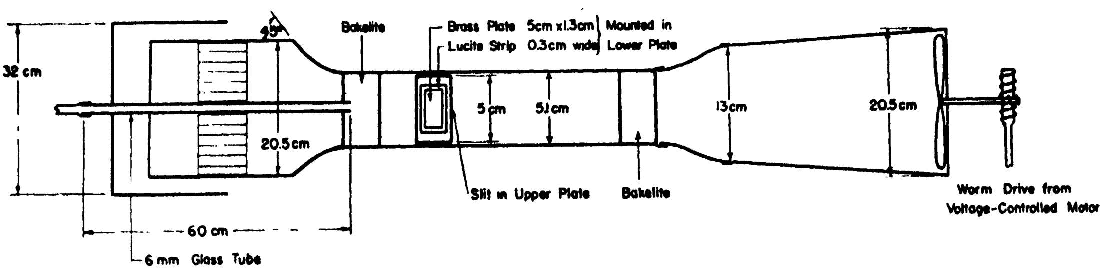
[5] H. Rohmann, Zeits. f. Physik 17, 253 (1923).



Perspective



Elevation



Plan

Not to Scale

FIG. 1. The precipitation apparatus.

parisons of experimental results with theoretical precipitation curves are presented here.

## APPARATUS

A diagram of the precipitation apparatus is shown in Fig. 1. The smoke, which was contained in a large flask, was forced by means of compressed air first through a short length of small-bore glass tubing (to reduce turbulence) and then into the entrance tube *T*. This tube extended through the entrance duct *D* to the rectangular precipitation chamber *C*. In order to minimize turbulence the entrance duct was provided with a honeycomb baffle *B* made from short lengths of Cellophane soda straws. Further reduction of turbulence was obtained by placing a box in front of the entrance duct in the position indicated in the plan and elevation of Fig. 1. The

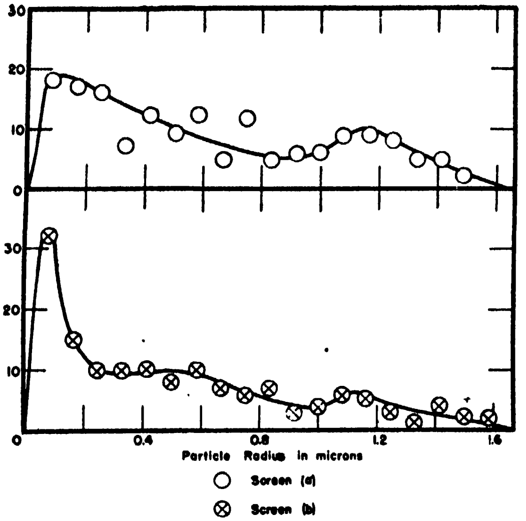


FIG. 2a. Size-frequency distributions of particles in a rosin smoke charged by means of a heated filament.

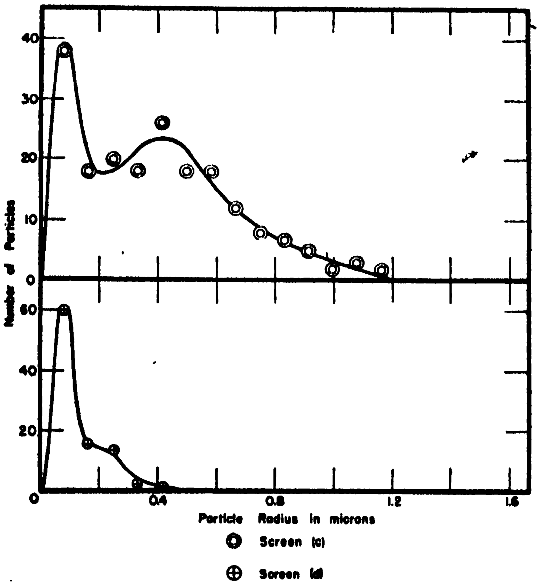


FIG. 2b. Size-frequency distributions of particles in a rosin smoke charged by means of a heated filament.

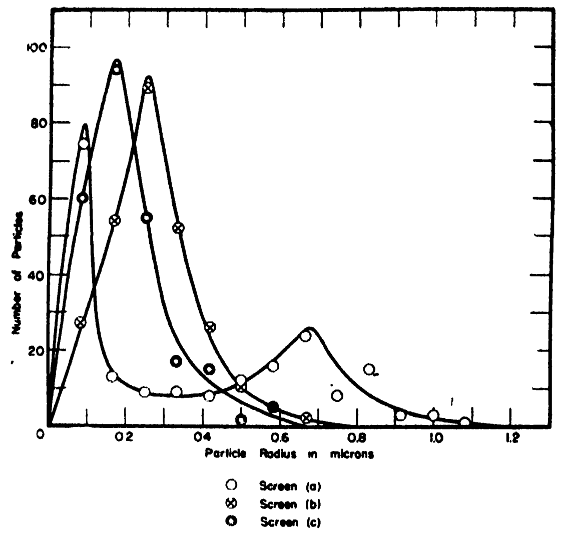


FIG. 3. Size-frequency distributions of particles in a rosin smoke charged by means of a cotton cloth at high potential.

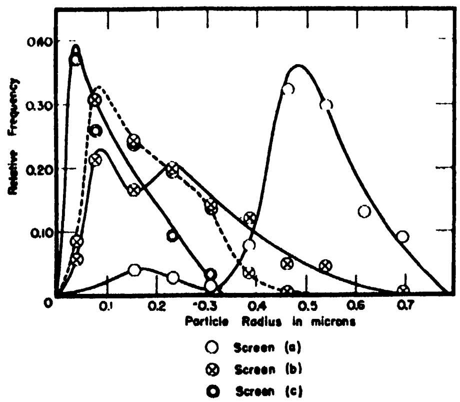


FIG. 4. Size-frequency distributions of particles in a rosin smoke charged by means of a plate coated with rayon flock fibers.

precipitation chamber consisted of two parallel brass plates *P* separated by lucite strips. The upper plate, which was charged to about −4 kv, was insulated from the entrance and exit ducts by bakelite strips *E*. The various charging devices which were tested were placed above or in the plane of an aperture *S* in the upper plate, and were surrounded by an air-tight insulator to prevent air from entering around the charging region. The lower plate was provided with a metallic collector *H* which was mounted in the plane of this plate and insulated from it. This collector was made movable along the direction of wind flow in order to measure, by means of a galvanometer, the iron currents emitted by the charging device as a function of the position of the collector. The air-stream outlet consisted of two sections which ended finally in a large circular cross-section. A four-bladed fan, run from a variable speed motor, was placed just inside the outlet section *N*; this fan produced air velocities at the center of the precipitation

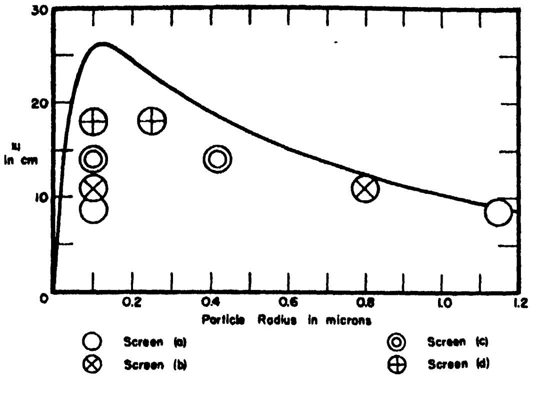


FIG. 5. Average particle radius as a function of point of precipitation; charging device was a heated filament.

chamber of 8 to 25 cm/sec. The velocities which were used in the runs described below were 12 to 15 cm/sec.

A uniform downward current of ions of about $10^{-8}$ ampere/cm$^2$ must be obtained in order to charge the smoke particles. Of the large number of devices which were tested for this purpose, the three most promising were investigated in detail

TABLE I. Summary of experimental results.

Run 1; Charging device: heated filament

| Screen | Particles measured | $x_i$ in cm | Position of maxima in microns |
|---|---|---|---|
| (a) | 176 | 8.7 | 0.1<br>1.15 |
| (b) | 146 | 11.0 | 0.1<br>0.5, 1.1<br>(Average, 0.8) |
| (c) | 195 | 14.0 | 0.1<br>0.42 |
| (d) | 99 | 18.0 | 0.1<br>0.25 |

Run 2; Charging device: cotton cloth

| Screen | Particles measured | $x_i$ in cm | Position of maxima and half width at half maximum, in microns |
|---|---|---|---|
| (a) | 196 | 6.5 | 0.08±0.05,<br>0.69±0.13 |
| (b) | 265 | 12.0 | 0.25±0.08 |
| (c) | 248 | 15.0 | 0.17±0.10 |

Run 3; Charging device: rayon flock fibers

| Screen | Particles measured | $x_i$ in cm | Position of maxima in microns |
|---|---|---|---|
| (a) | 77 | 4.0 | 0.16<br>0.48 |
| (b) | 229 | 6.0 | 0.08<br>0.23 |
|  | 452 | 6.0 | 0.09<br>(0.2 to 0.3) |
| (c) | 97 | 8.6 | 0.05 |

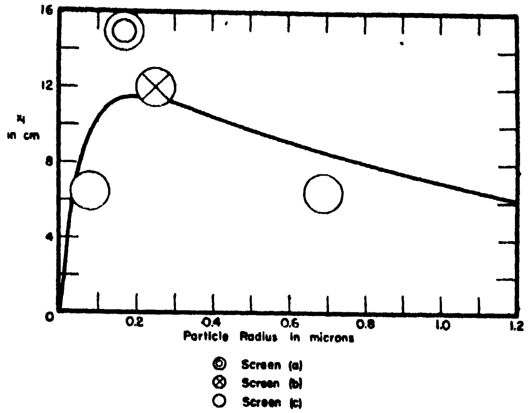


FIG. 6. Average particle radius as a function of point of precipitation; charging device was a cotton cloth.

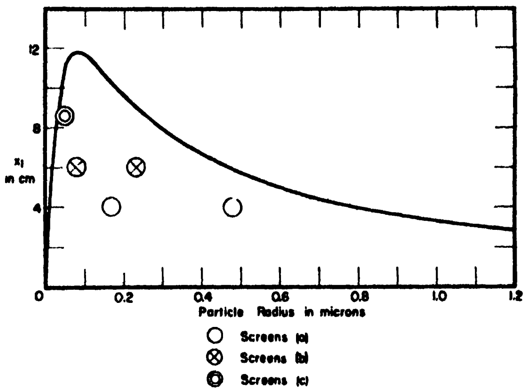


FIG. 7. Average particle radius as a function of point of precipitation; charging device was made from rayon flock fibers.

by studies of the size-frequency distribution of the smoke particles precipitated at various points in the apparatus. The first (Run 1) consisted of a small heated Nichrome coil which was supported by a short quartz rod and jacketed by a Pyrex compressed air cooler. Immediately below this coiled filament was placed a wire screen (15–50 meshes/in.$^2$) which was allowed to remain at "floating" potential. The filament was heated by means of a filament transformer and was maintained at a potential of about −5 kv. The second charging device (Run 2) consisted of a small area (1 cm×2.5 cm) of cotton bird's-eye toweling which was placed in contact with a metal plate at −4 to −10 kv. The third charging device (Run 3) consisted of a small aluminum plate maintained at −14 kv and covered sparsely with rayon flock fibers (about 250/cm$^2$) attached normal to it.[6] This third charging device gave the most nearly reproducible and uniform charging conditions of all of those which were tested.

For studies with the electron microscope, collodion-covered screens were placed on the lower plate of the precipitation apparatus at various distances from the charging region. These screens were 6-mm circles cut from 350-mesh brass screen, and were coated with a collodion film about 0.1 micron in thickness.

## EXPERIMENTAL PROCEDURE

The wind velocity $v_0$ was measured in the following manner: a thin stream of smoke was allowed to flow down the center of the precipitation chamber. A disturbance was given to the stream by a slight movement of the inlet tube, and the time required for this disturbance to move a known distance was measured.

The ion current in the charging region was measured at various positions by means of the movable collector plate; it was found to be fairly uniform and limited to a small region. The extent of the charging region, extrapolated to the height of the smoke stream above the lower plate, divided by the velocity of the stream was taken as the time of charging, $t$. The average ion current per unit area in the charging region was taken as the ion current density, $i$.

The rosin smokes were prepared by heating about 70 mg of rosin in a small copper pan; the rosin condensed into a smoke consisting of approximately spherical particles of radius $0.05 < a < 3$ microns. The smoke was collected in the large flask and was forced into the apparatus by a slow stream of compressed air. When the charging and deflecting potentials were turned on, the smoke stream was observed to curve down slightly and fan out into a curtain of smoke dropping to the lower plate. The smoke thus deposited a narrow track of rosin on the lower plate where, for the purpose of electron microscopy, the small collodion-covered collectors were placed.

Photographs were taken of the smoke deposits on the collodion-covered screens with the use of the electron microscope designed at the California Institute of Technology by Professor W. V. Houston and Dr. J. H. Bradner, and the RCA electron microscope at the Department of Physics of the University of California at Berkeley. Enlargements of the electron microscope photographs were made to a total linear magnification of about 6000. For spherical particles the diameter was measured; for elliptical and slightly irregular particles the average of the largest and smallest diameters was taken. Nearly all of the rosin smoke particles were approximately spherical except for the larger particles obtained in Run 1, in which the heated filament was used as a charging device. These highly irregular particles were divided into spherical or approximately elliptical sections and measured as independent parts. Although this idealization is clearly not a good approximation, a convenient method of comparing particle sizes was obtained by this method. There is no satisfactory treatment of the law of fall for particles of irregular shape. From the measurements of particle size, the size-frequency distribution curves were prepared by plotting the number of particles of a given size as ordinate against the radius of the particle in microns as abscissa, and then drawing a smooth curve through the points.

## RESULTS

The size-frequency distribution curves obtained in the three runs are shown in Figs. 2a, 2b, 3, and 4. The experimental results are summarized in Table I. On Screen (b) of Run 1 two maxima appeared for $a > 0.3$ microns whereas the theory required only the one shown by most of the other experimental results. The average of these two peaks was taken. It is to be noted that the numbers of particles of given size determining parts of this and some of the other distribution curves are too small to yield satisfactory statistical distributions. As a result the method of interpolation by which some parts of the curves should be obtained is not unambiguous. The photographed area of Screen (b) of Run 3 was divided into two parts, and the particles on these two parts were counted separately in order to determine whether the result-

---

[6] Flock is a preparation of textile fibers cut to uniform, short lengths. In order to prepare this charging device the required density of flock was placed on a grounded conducting plate. The aluminum plate, wet with Glyptal, was suspended above the grounded plate and charged to about −15 kv in order to deposit the fibers in an erect position on the charged plate. The potential was left on until the Glyptal hardened.

ing size-frequency distribution curves were in satisfactory agreement. The results, shown in Fig. 4 and Table I are in good agreement, but it is quite clear that a large number of particles must be counted before satisfactory distributions are obtained.

The experimental results are compared with the theoretical curves, calculated as described in the Appendix, in Figs. 5, 6, and 7. The numerical expressions from which the curves were calculated are, for Run 1,

$$x_i = \frac{1660a_\mu}{(1+0.082/a_\mu)Q} \text{ where } Q = 123a_\mu{}^2 + 23a_\mu;$$

for Run 2,

$$x_i = \frac{1840a_\mu}{(1+0.082/a_\mu)Q} \text{ where } Q = 157a_\mu{}^2 + 82a_\mu;$$

and, for Run 3,

$$x_i = \frac{1840a_\mu}{(1+0.082/a_\mu)Q} \text{ where } Q = 476a_\mu{}^2 + 39a_\mu,$$

where $a_\mu$ is the particle radius in microns. The agreement between the curves and the experimental points is not too good. However, the general character of the theoretical curves is reproduced by the experiments, for example, the precipitation of both large and small particles at a given point as required by the theory appears to be verified.

In order to test whether the original smoke sample itself contained an abnormally large number of small particles, the whole thread of smoke was swept repeatedly across a collodion-covered screen by variation of the corona potential, until a satisfactory density of particles was obtained. Electron microscope photographs of this sample did not show a predominant number of particles of radius less than 0.2 micron; the size-frequency distribution curve for the unseparated smoke showed a single maximum at about 1 micron.

Considerable experimental difficulties were encountered in the use of the various charging devices. The filament heated the charging region considerably, and even though approximate thermal equilibrium had been reached before the run was started some slight thermal unsteadiness in the precipitation conditions persisted. The cotton cloth charging device produced results which were quite unreproducible; however, no variations in the charging conditions were usually observed during a single run. The conducting plate covered with the rayon flock fibers was probably the most satisfactory charging device which was tested. These fibers undoubtedly retained their relative orientation for successive runs better than did the cloth fibers. The appearance of the track of rosin precipitated on the lower plate was very similar for successive runs with the flock fiber charging device, although a series of five successive runs under as nearly the same conditions as possible gave results which agreed rather poorly.

The reasons for this poor agreement probably include (a) variations in the initial charges on the particles, (b) turbulence in the wind flow in the tunnel, (c) irregularities in the particle shape, (d) fluctuations in the charging current, and (e) statistical variations in the charging. The effects of variation of the initial charges of the particles were estimated by measurement of the maximum deflection of the smoke stream in a uniform field in the absence of the precipitating current. The maximum deflection observed corresponded to a maximum initial charge of $\pm 8e$ on a particle of radius 0.5 micron, or $\pm 34e$ on a particle of radius 2 microns. Turbulence is extremely difficult to eliminate completely, and its effect was apparent on careful visual observation under the best conditions which we obtained; it may have been caused, in part, by electrical wind effects. Fluctuations in the charging current of 1 to 2 percent were detected with a probe which was connected to an electrometer circuit. The effect of irregularities in the shapes of the particles and of statistical variations in charging could not be estimated.

## ACKNOWLEDGMENT

Our thanks are due Professor Linus Pauling, who suggested this study, for his interest, advice, and encouragement. We are also indebted to Mr. George Standart for assistance with part of the experimental work, and to Dr. R. V. Langmuir and Mr. R. Rau of the California Institute of Technology, Mr. A. E. Smith of the Shell Development Company, and Mr. A. W. Einarsson of the University of California at Berkeley

for their assistance in helping us to obtain electron microscope photographs.

## APPENDIX: THEORETICAL SECTION

A particle which has a radius $a$ and charge $Qe$ and is swept along a horizontal rectangular channel (see Fig. 8) between two charged plates has a downward velocity $v'$ because of the gravitational and electrostatic forces, $mg+E_0Qe$;

$$v'=\frac{mg+E_0Qe}{6\pi\eta a}\left(1+\frac{A\lambda}{a}\right), \tag{1}$$

where $\eta$ is the viscosity of the gas; $A$ is the Cunningham constant[7]; and $\lambda$ is the mean free path of the gas molecules.

The problem of the gas velocity at any point in a tube of rectangular cross section has been solved by Cornish.[8] In a tube with cross-sectional width $c$ and height $d$ in which the velocity on the axis ($y=z=0$) is $v_0$, the velocity of gas flow along any line parallel to the axis is

$$v=v_0\frac{\left[\frac{1}{4}-\frac{y^2}{d^2}-\frac{8}{\pi^3}\left(\frac{\cosh\frac{\pi z}{d}}{\cosh\frac{\pi c}{2d}}\cos\frac{\pi y}{d}-\frac{1}{3^3}\frac{\cosh\frac{3\pi z}{d}}{\cosh\frac{3\pi c}{2d}}\cos\frac{3\pi y}{d}+\cdots\right)\right]}{\frac{1}{4}-\frac{8}{\pi^3}\left(\frac{1}{\cosh\frac{\pi c}{2d}}-\frac{1}{3^3\cosh\frac{3\pi c}{2d}}+\cdots\right)} \tag{2}$$

The horizontal distance, $x_i$, from the point at which the particles are charged to the point at which they strike the lower plate may then be shown, by integration of the equation $dx=(v/v')dy$, to be

$$x_i=\frac{6\pi\eta a v_0}{(mg+E_0Qe)\left(1+\frac{A\lambda}{a}\right)}\frac{2l^2}{d}\left(1-\frac{2l}{3d}\right)$$
$$\times\left\{1+\frac{32}{\pi^3\cosh\frac{\pi c}{2d}}\left[1+\frac{32}{\pi^3\cosh\frac{\pi c}{2d}}\right]\right.$$
$$\left.\times\left[1-\frac{d^2\left(1-\cos\frac{\pi l}{d}\right)\cosh\frac{\pi z}{d}}{2\pi l^2\left(1-\frac{2l}{3d}\right)}\right]+\cdots\right\}, \tag{3}$$

where $l$ is the height above the lower plate at which the stream of particles is charged. For the conditions under which the experiments were carried out $E_0Qe\gg mg$ and $z=0$, and an adequate approximation to Eq. (3) is

$$x_i=\frac{6\pi\eta a v_0}{E_0Qe\left(1+\frac{A\lambda}{a}\right)}\frac{2l^2}{d}\left(1-\frac{2l}{3d}\right). \tag{4}$$

This equation would be suitable for calculation if we knew the number of charges, $Q$, as a function of the particle radius and known or measurable quantities.

A theory of the charging of a spherical particle of radius $a$ and dielectric constant $K$ passing through a downward current of ions has been derived by Rohmann[8] under the assumptions that (1) diffusion of charges to the particle is negligible, (2) electrical image forces between an ion approaching a particle and the particle are negligible, and (3) all of the incident ions give up their charge to the particle. His expression for the charge $Q_fe$ acquired in time $t$ by an initially uncharged particle may be written in the form

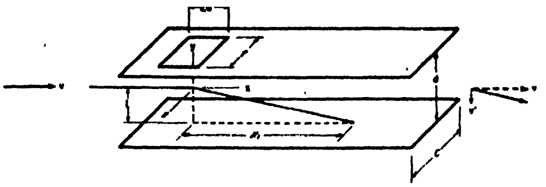


FIG. 8.

$$Q_fe=\left(1+\frac{2(K-1)}{K+2}\right)E_1a^2\frac{\pi it}{E_1+\pi it}, \tag{5}$$

where $i$ is the ion current per unit area and $E_1$ is the field intensity in the charged region.

No satisfactory theory has been developed for the initial stage of the charging of a spherical particle of low dielectric constant by diffusion. The final stage of diffusion charging was investigated theoretically and experimentally by Arendt and Kallmann.[9] Their work, and also the experimental work of Deutsch[10] and Schweitzer[11] showed that the saturation charge taken up by oil drops of radius from 0.1 micron to 10 microns in a space charge of the order of $10^6$ charges/cm$^3$ is proportional to the particle radius. Ladenburg[12] pointed out that in the process of precipitation of smoke particles in a charging field there are two distinct charging mechanisms, one, important for the larger particles, giving a charge proportional to $a^2$ as described in the preceding paragraph, and another, important for the smaller particles, giving a charge proportional to $a$. This second mechanism is called diffusion charging although it includes, in addition to ordinary diffusion of ions to the particle, the effect of image forces on the charge approaching the dielectric surface of the particle and the effect of the net charge on the diffusion of ions to the particle. We have assumed that the number of charges diffusing to a particle of radius $a$ in a time $t$ and in an ion atmosphere of volume density $f$ is

$$Q_d=Cfta. \tag{6}$$

The value of the proportionality constant was determined to be $C=0.07$ by extrapolation to zero time of the experimental curves for the charging of oil drops as a function of time, as given by Arendt and Kallmann.[9] Unfortunately, Arendt and Kallmann did not obtain data for times less than about 4 min., and consequently the constant $C$ cannot *a priori* be regarded as determined even to the correct order of magnitude for the time intervals of interest ($t\sim0.2$ sec.). We shall, nevertheless, use the resulting expression for calculation of $Q_d$ because our results described above indicate that the resulting term in the

---

[7] E. Cunningham, Proc. Roy. Soc. **A83**, 357 (1910).

[8] R. J. Cornish, Proc. Roy. Soc. **A120**, 691 (1928).

[9] P. Arendt and H. Kallmann, Zeits. f. Physik **35**, 421 (1926).

[10] W. Deutsch, Zeits. f. tech. Physik **7**, 623 (1926).

[11] H. Schweitzer, Ann. d. Physik [5] **4**, 33 (1930).

[12] R. Ladenburg, Ann. der Physik [5] **4**, 863 (1930).

charging law is approximately correct, and is almost certainly of the correct order of magnitude. Since the ion density is given by the relation $f=i/(E_1ue)$, where $u$ is the mobility of the ions, Eq. (6) therefore becomes

$$Q_de=0.07iat/E_1u. \tag{7}$$

In the absence of a complete mathematical treatment of the charging process, we shall write the expression for the total charge at a time $t$ on a spherical particle of radius $a$ in the form

$$Qe=Q_fe+Q_de, \tag{8}$$

where $Q_fe$ and $Q_de$ are given by Eqs. (5) and (7), respectively. In this and the preceding equations the effect of the comparatively small initial charge on the particle has been omitted. Although the two charging processes are probably not independent, and the expressions for the two processes therefore not strictly additive, it may be pointed out that because of the different dependence of the two processes on the particle radius, Eq. (7) represents the important term for the smaller particles, whereas Eq. (6) represents the important term for the larger particles.

The value of $Qe$ from Eq. (8) may be substituted into Eq. (4) in order to calculate the distance along the direction of gas flow from the center of the charging region to the point of precipitation of a particle of radius $a$. In this calculation we have made the additional assumption that the charging, which occurs over the distance $v_0t=2a'$ (Fig. 8), may be approximated by complete charging at a point at the center of the charging region.

In some of the experiments the field in the charging region was considerably higher than that between the deflecting plates of the rectangular precipitation chamber. The average field in the charging region in this case was calculated from an approximate solution of the two-dimensional potential distribution by expansion in rectangular harmonics. For the case (Run 3, above) in which this treatment was applied, the field thus calculated was about 7 percent smaller than that obtained from the relation $E_1=-V_1/d$, where $V_1$ is the potential of the upper plate of the rectangular tube. The effect of space charge in decreasing the field intensity in the charging region was small enough to be neglected.

---

# Dislocation Theory as Applied by N.A.C.A. to the Creep of Metals

A. S. Nowick and E. S. Machlin
*Aircraft Engine Research Laboratory, Cleveland, Ohio*

(Received September 9, 1946)

An equation for the steady-state rate of creep of pure annealed polycrystalline metals is derived through dislocation theory and the theory of rate processes. The rate of generation of dislocations is shown to be the rate-determining process. A specific mechanism for the generation of a dislocation is presented, from which the heat and entropy of activation of the process are expressed in terms of physical constants of the material. In addition the lowering under stress of the potential energy barrier, which yields the stress-dependent term, is given in terms of constants of the material and a "back-stress" term. The creep equation thus obtained is found to be in good agreement with data in the literature.

## INTRODUCTION

Creep of metals is generally subdivided into three stages, as shown in Fig. 1. The initial stage, in which the slope of the elongation-*versus*-time curve (the rate of creep) is rapidly decreasing, is commonly designated the primary or transient stage; secondary or steady-state creep refers to the straight-line portion of the creep curve. After a sufficient length of time, the rate of creep increases in the region designated the tertiary stage.

The criterion generally used to evaluate the creep property of a given material at a particular stress and temperature is the steady-state creep rate. One of the first attempts to analyze steady-state creep by other than empirical methods was by Kanter.[1] Later, Kauzmann[2] applied Eyring's theory of liquid flow, which is a special application of the theory of rate processes, to the steady-state flow of metals. The problem of creep has also been attacked by the use of the theory of dislocations. A survey of results obtained thus far by this approach is reported by Seitz and Read.[3] Although these three treatments predict equations for the steady-state creep rate that show the correct dependence on stress and temperature, quantitative predictions cannot be

---

[1] J. J. Kanter, Trans. A.I.M.E. **131**, 385 (1938).

[2] W. Kauzmann, Trans. A.I.M.E. **143**, 57 (1941).

[3] F. Seitz and T. A. Read, J. App. Phys. **12**, 100, 170, 470, 538 (1941).

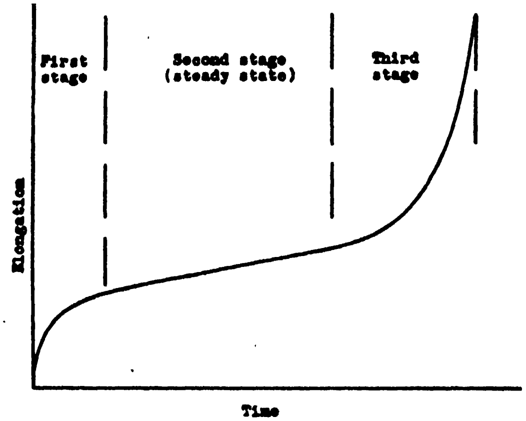


FIG. 1. A typical creep curve showing the three stages.

made from them as to how physical and structural constants of materials affect creep rate. Furthermore, difficulties were encountered in attempting to explain the large negative entropy of activation that experimental data showed.

This report derives an equation for the steady-state creep rate in terms of a specific dislocation mechanism which involves considering the generation of dislocations to be a rate process. The features of the theory of rate processes and of dislocation theory required for the development of the creep equation are first briefly reviewed; the derivation then follows.

The material contained herein was derived at the Cleveland laboratory of the N.A.C.A. and was originally presented as N.A.C.A. Technical Note No. 1039.

## THEORY OF RATE PROCESSES

The theory of rate processes,[4] as developed by Eyring and others, considers a reaction or any rate process to be the result of the crossing of a potential-energy barrier by molecules whose energies have exceeded a certain minimum value. The reaction consists of the formation of an "activated complex" capable of crossing the barrier followed by the passage of this complex over the barrier. The most important assumption is that the initial reactants and activated complexes are always in equilibrium. Application of thermodynamic and statistical mechanical considerations shows that the number $r$ of activated complexes passing over the barrier per second (or rate of reaction), per unit concentration of reactants is given by

$$r=\frac{kT}{h}\exp(-\Delta F_a/kT) \qquad (1)$$

in terms of Boltzmann's constant $k$, Planck's constant $h$, the absolute temperature $T$, and the free energy of activation per molecule, $\Delta F_a$.

The term $kT/h$ can be regarded as the effective frequency at which activated complexes cross over the barrier; the exponential factor represents the probability of formation of an activated complex. In the calculation of this probability term, the contribution due to the translational degree of freedom along the "reaction coordinate" (the most favorable reaction path on the potential-energy surface) has been disregarded because it is included in the factor $kT/h$. The free energy of activation $\Delta F_a$ is interpreted as an ordinary free-energy term and can be expressed by

$$\Delta F_a=\Delta H_a-T\Delta S_a, \qquad (2)$$

where the heat of activation per molecule $\Delta H_a$ is the height of the potential-energy barrier, and $\Delta S_a$ is the entropy of activation per molecule.

## DISLOCATION THEORY IN RELATION TO SLIP

Of several mechanisms suggested, the one that has been most successful in explaining the slip process is the theory of dislocations. This theory proposes that local deviations from a perfect lattice, called dislocations, exist in single crystals and that movement of dislocations through the stressed crystal produces slip. A dislocation consists in a stable arrangement of atoms such that, in a region of a few atomic distances, $n+1$ atoms in the slip direction face an adjacent parallel row of $n$ atoms across the slip plane. The point at which the atoms are one-half an atomic spacing "out of step" is called the center of the dislocation. The extension of the dislocation in the slip plane and normal to the slip direction is called the length of the dislocation. Two types of dislocations can exist: a dislocation is called positive if in the neighborhood of the center of the dislocation the compression is above the slip plane, and negative if the compression occurs

---

[4] S. Glasstone, K. J. Laidler, and H. Eyring, *The Theory of Rate Processes* (McGraw-Hill Book Company, Inc., New York and London, 1941), Chaps. I, IV, and IX.

below the slip plane. Positive and negative dislocations move in opposite directions under a given shear stress, but they produce the same net result in passing through the crystal, a relative translation of one atomic distance between the parts of the lattice above and below the slip plane.

Inasmuch as the passing of a dislocation out of the crystal lattice results in its loss, new dislocations must be generated in order for continuous plastic deformation to be possible. Because of the high energy required, the process generally takes place at regions of high stress concentration resulting from imperfections in the single crystals. The existence of such imperfections as a result of accidents of growth during the formation of single crystals is generally accepted, but the nature of the flaws that exist is still a matter of controversy. For example, the theory that single crystals are made up of a mosaic block structure is discussed by Seitz and Read. X-ray evidence indicates that the average spacing of the imperfections is of the order of 1 micron. The dynamic dislocation picture obtained is as follows: dislocations are generated under shear stress at the regions of high stress concentration, which will be called sources of dislocations. Upon generation, the dislocations are only a few atoms long because of the extremely low probability of generating a full-length dislocation (one whose length is of the order of the spacing between imperfections). As the dislocations move through the crystal under a shear stress, they increase in length and move in the slip plane and slip direction until they become stuck at an imperfection or in the vicinity of other dislocations.

Further details on the properties of dislocations as used to explain the qualitative aspects of plastic deformation are given by Seitz and Read.[3]

### DERIVATION OF THE CREEP EQUATION

In the analysis of the creep process, it will be assumed that creep as well as slip takes place by the motion of dislocations. If $L$ is the spacing between imperfections (in terms of the mosaic structure theory, the length of a mosaic block), we choose a cubic block of side $L$ and make two assumptions: (a) no "annihilation" (union of pairs of dislocations of opposite sign) takes place so that each dislocation passes completely through the block; (b) the block chosen is representative of the entire crystal in that it will shear at the same rate as any other similar block. This assumption follows from the requirement of a steady-state condition.

For the chosen block, the number of dislocations generated per second within the block must equal the number leaving the block per second, or

$$R_g N_s L^3 = N_d v L, \tag{3}$$

where $R_g$ is the rate of generation for a single generating source; $N_s$ is the number of sources per unit volume; $N_d$ is the number of dislocations per unit area that intersect a plane normal to the slip plane and containing the slip direction; and $v$ is the average velocity of the moving dislocations. Inasmuch as the displacement produced when a dislocation passes out of the block is equal to the lattice spacing $d_1$ between atoms in the slip direction, the shear rate $u'$ is given by

$$u' = \frac{1}{L}(N_d v L) d_1 = N_d v d_1. \tag{4}$$

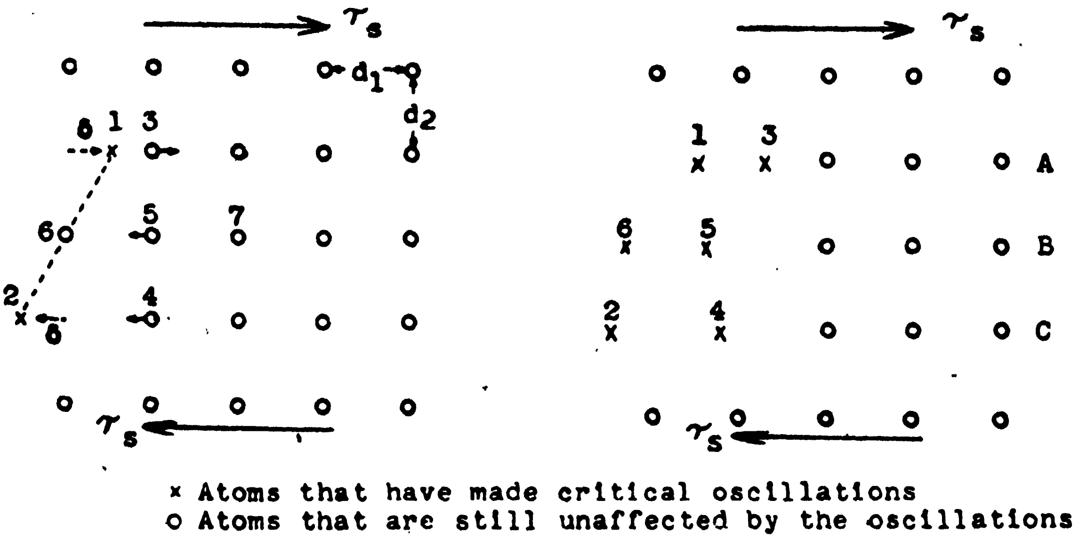
The occurrence of slip bands at distances apart of the order of 1 micron indicates an average of about one source of dislocations for each block so that

$$u' = (d_1/L) R_g. \tag{5}$$

Up to this point, only creep of single crystals has been considered. In polycrystalline metals, it would appear that two types of creep are possible: creep within individual grains, and creep resulting from intergranular motion. As noted by Seitz and Read, present data on the effect of grain size on creep leave doubt as to the possibility of the occurrence of intergranular creep. For example, it has been shown[5] in work on copper of various grain sizes that there are no detectable differences in creep strength other than those resulting from oxidation. (Unfortunately this work was carried out only at a single temperature.) Creep in polycrystalline metals is therefore assumed to be predominately the result of deformation occurring within the individual grains by the motion of dislocations. If the

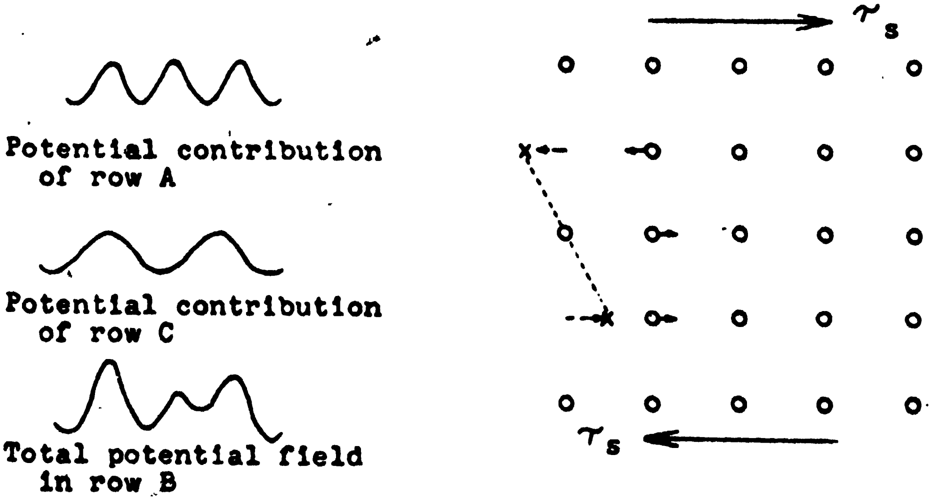
---

[5] E. R. Parker and C. F. Riisness, Metals Tech. A.I.M.E. 11 (1944), Tech. Pub. No. 1690, p. 1.


τs
d1
d2
1
3
5
7
6
2
4
A
B
C
τs
× Atoms that have made critical oscillations
o Atoms that are still unaffected by the oscillations

(a) Activated complex for a positive dislocation.

(b) Dislocation formed by rearrangement of the activated complex.


Potential contribution of row A
Potential contribution of row C
Total potential field in row B
τs
τs

(c) Determination of potential field in row B.

(d) Activated complex for a negative dislocation

FIG. 2. Generation of a dislocation at a positive source.

orientation of the grains is such that slip occurs in the plane of maximum shear stress, then, following Kauzmann, the strain rate $u'$ can be corrected to the tensile creep rate $u$ by an additional factor of about $\frac{2}{3}$. Such a factor would be of no consequence here because $d_1/L$ is known only in order of magnitude; accurate values cannot be given until more is known about the nature of the crystal imperfections. The tensile creep rate $u$ of a polycrystalline material can therefore be written approximately as

$$u=(d_1/L)R_g. \qquad (5')$$

In calculating the rate of generation $R_g$, the generated dislocation will be taken to be 1 atom long. A simple model for the generation of a positive dislocation at a right-hand source[6] is presented in Figs. 2(a) and (b). If atoms 1 and 2 are simultaneously displaced in the directions shown in Fig. 2(a) by large fractions of the lattice spacing, and atoms 3, 4, and 5 are moving in the slip direction as shown, the generation of a dislocation becomes possible. Before a dislocation will form, however, a rearrangement of the neighboring atoms must take place in order to achieve a stable configuration. Figure 2(b) shows the rearranged condition. The new positions of atoms 5 and 6 after rearrangement were found by using the approximate method of Taylor[7] in his discussion of the motion of a dislocation under stress. The contributions to the potential field in row $B$ by the neighboring rows $A$ and $C$ are taken to be sinusoidal with periods equal to the spacing of atoms 1 and 3 in row $A$ and atoms 2 and 4 in row $C$. Figure 2(c) shows that addition of the two sinusoidal potential functions gives a resultant potential function having four minima. Atoms 5, 6, and 7 can be expected to fall into the deepest of these minima if atom 5 had begun by oscillating to the left as in Fig. 2(a).

The displacement $\delta$ of atoms 1 and 2 will be some fraction $f$ of the atomic separation $d_1$ in the slip direction. A value of $f$ of somewhat less than $\frac{1}{2}$ makes it possible to form a stable configuration with a minimum activation energy.

The mechanism just presented is an example of a rate process where the "reactant" is the perfect lattice at the region of high stress concentration. The "activated complex" for a positive dislocation, which for brevity will be called a positive activated complex, can be taken as the configuration shown in Fig. 2(a) and the "product" as the dislocation after atomic rearrangement has taken place (Fig. 2(b)). If the fundamental assumption of rate-process theory that equilibrium exists between reactant and activated complex is made, the considerations involved in the theory of rate processes can be applied to the generation of dislocations. If $t$ is the average time for the activated complex to pass into final form, the rate of generation of positive dislocations $R_g^+$ can be expressed as

$$R_g^+=(1/t)P^+, \qquad (6)$$

where $P^+$ is the probability of formation of a positive activated complex at a right-hand source.

It has been shown[8] that if the displacement involved in the transition from the activated complex to the final state is on the order of $10^{-8}$ centimeter, the factor $1/t$ is approximately equal to $kT/h$. In the formation of a dislocation, this transition corresponds to passing from the configuration of Fig. 2(a) to that of Fig. 2(b), which clearly involves a displacement of this order of magnitude. Hence, $R_g^+$ is given approximately by

$$R_g^+=(kT/h)P^+. \qquad (7)$$

Similarly, for the rate of generation of negative dislocations,

$$R_g^-=(kT/h)P^-, \qquad (7')$$

where $P^-$ is the probability of formation of a negative activated complex at a right-hand source.

In order to simplify the calculation of $P^+$ and $P^-$, the generation of a dislocation will be treated as a "local shear" which takes place by means of thermal oscillations. The energy involved in this process will be calculated by the theory of elasticity. Furthermore, Hooke's law will be assumed valid for large strains. The use of these assumptions is necessary because the forces between individual atoms have not been evaluated in terms of physical constants, but it should be remembered that the approximations may be serious enough to lead to large errors in the results. The thermal atomic oscillations in Fig. 2(a) will be regarded as resulting in a local shearing of atoms 1, 6, and 2 producing a strain of $\delta/d_2$. Inasmuch as the elastic energy per unit volume for a shear stress $\tau$ is $\tau^2/2G$ where $G$ is the modulus of rigidity, and the volume of the

---

[6] A source of dislocations will be called a right-hand source if it generates dislocations to the right, and a left-hand source if it generates to the left. The sources are taken to be at "internal surfaces," for example, the edge of a surface of misfit between mosaic blocks.

[7] G. I. Taylor, Proc. Roy. Soc. (London) **A145**, 362 (1934).

[8] Glasstone, Laidler, and Eyring, reference 4, p. 189.

region of shear under consideration is twice the atomic volume $V$, the elastic energy involved is

$$W=V\tau^2/G. \tag{8}$$

Now, let $\tau_0$ be the local shear stress attained by atoms 1, 6, and 2 in Fig. 2(a). If the displacement $\delta$ is some fraction $f$ of $d_1$, application of Hooke's law to the displacement yields

$$\tau_0=Gxf, \tag{9}$$

where $x$ is defined as the ratio $d_1/d_2$. If the source of generation is under a stress $\tau_s$, as shown in Fig. 2, the energy $W_t$ to be attained thermally is, by (8)

$$W_t=(V/G)(\tau_0-\tau_s)^2. \tag{10}$$

Inasmuch as $\tau_s$ may be expected to be only a small fraction of $\tau_0$ for rigid materials, the term in $\tau_s^2$ may be neglected in the binomial expansion of $(\tau_0-\tau_s)^2$, so that, incorporating (9), we obtain

$$W_t=VGx^2f^2-2Vxf\tau_s. \tag{11}$$

In order to evaluate $\tau_s$, work-hardening must be considered. A crystal in the steady-state range of creep is generally work-hardened because deformation has taken place. From the nature of a dislocation it is clear that a stress field exists about it. By making use of this stress field, Taylor[7] has explained work-hardening on the basis of a "lattice" of stuck dislocations. Taylor's explanation is a static one and is applied only to determining the form of the stress-strain curve. In terms of a dynamic approach, the existence of a stress field about a dislocation leads to the concept suggested by Kochendörfer[9] that the lattice of stuck dislocations creates an inner stress field whose direction at the point of generation is opposite to the externally applied shear stress $\tau_e$ in the slip plane. The value of the inner stress field at the point of generation will be called the back stress $\tau_b$. The dynamic concept of hardening therefore consists in the formation of a lattice of stuck dislocations, which results in a back stress at the point of generation. This back stress lowers the effective shear stress at these points, and as a result the rate of generation, hence the creep rate, decreases. These considerations lead to the result that

$$\tau_s=q(\tau_e-\tau_b), \tag{12}$$

where $q$ is the stress concentration at the generating source. Inasmuch as $\tau_b$ is always less than $\tau_e$, the directions of $\tau_e$ and $\tau_s$ are the same.

In addition to the energy of local shear, the configuration of Fig. 2(a) involves a limitation of the movement of atoms 1, 2, 3, 4, and 5 to a particular crystallographic direction within a very small solid angle. If the probability of oscillation in this direction is $p$, the total probability of the configuration of Fig. 2(a) is

$$\begin{aligned}P^+&=p^5\exp(-W_t/kT)\\&=p^5\exp[-(VGx^2f^2-2Vxf\tau_s)/kT].\end{aligned} \tag{13}$$

Similarly the probability of forming a negative activated complex (the configuration of Fig. 2(d)) at a positive source is

$$P^-=p^5\exp[-(VGx^2f^2+2Vxf\tau_s)/kT], \tag{13'}$$

because $\tau_s$ is opposite to $\tau_0$ in this case. The applied stress $\tau_e$ therefore makes the generation of a positive dislocation more probable than the generation of a negative dislocation at a right-hand source.

If a negative dislocation is generated in the region under discussion, it would move to the left and out of the crystal without producing deformation if it were the only dislocation present. The last positive dislocation generated from this source, however, became stuck after moving through part of the crystal lattice; that is, all forces upon it were balanced. The generation of a negative dislocation at the point being considered will upset this equilibrium and attract the previous positive dislocation. As the two dislocations approach each other, the stress that each exerts on the other increases,[10] and a union takes place resulting in the annihilation of the two dislocations and restoration of a perfect lattice. The net rate of generation $R_g$ of positive dislocations that will eventually produce plastic deformation is $R_g^+-R_g^-$ or from Eqs. (7), (7′), (12), (13), and (13′)

$$\begin{aligned}R_g=\frac{2kT}{h}&\exp[-(VGx^2f^2+CT)/kT]\\&\times\sinh[2qVxf(\tau_e-\tau_b)/kT],\end{aligned} \tag{14}$$

where

$$C=-k\ln p^5. \tag{15}$$

---

[9] A. Kochendörfer, Zeits. f. Physik 108, 244 (1938).

[10] From the equation for the stress field about a dislocation; see Taylor, reference 7.

The equation for the net rate of generation of negative dislocations at a negative source is clearly identical with (14).

According to this derivation, it is now clear that the generation of a dislocation 2 atoms in length involves the simultaneous oscillation of twice as many atoms with sufficiently great energies. Thus, the probability of such a generation taking place can be neglected as compared with the probability of generating a 1-atom dislocation.

Comparison of (14) with Eqs. (1) and (2) now shows that the heat of activation (the height of the barrier for zero applied stress) and the entropy of activation are given, respectively, by

$$\Delta H_a = VGx^2f^2, \tag{16}$$

$$\Delta S_a = -C. \tag{17}$$

$C$ is always positive because $p$ is less than 1, so $\Delta S_a$ is negative.

When the expression (14) for $R_a$ is combined with the creep equation (5′), the complete equation for the steady-state creep rate $u$ becomes

$$u = \frac{2d_1}{L}\frac{kT}{h}\exp[-(VGx^2f^2+CT)/kT] \times \sinh[qVxf(\sigma-2\tau_b)/kT], \tag{18}$$

where the applied tensile stress $\sigma$ has been substituted for $2\tau_e$ assuming that slip in the individual grains occurs in the plane of maximum shear stress. For high values of $\sigma$, the hyperbolic sine function becomes an exponential, or in logarithmic form

$$\ln u = \ln\left(\frac{d_1kT}{Lh}\right) - \left(\frac{VGx^2f^2+CT}{kT}\right) + \frac{qVxf}{kT}(\sigma-2\tau_b). \tag{19}$$

## COMPARISON OF THEORETICAL AND EXPERIMENTAL CREEP EQUATIONS

The evaluation of the back stress $\tau_b$ as a function of $T$ and $\sigma$ is difficult and has not yet been accomplished. We will therefore compare the theoretical results with empirical equations. It has been found that a hyperbolic sine function represents the dependence of creep rate on applied tensile stress $\sigma$ for many polycrystalline materials[11] in agreement with Eq. (18). It has also been shown[12] that creep rates of a large number of materials can be represented in the form

$$\ln\frac{u}{T} = -\frac{a}{T} - b + \gamma\sigma, \tag{20}$$

where $a$ and $b$ are constants for a given material and $\gamma$ is a function of temperature but not of stress. If the notation is changed, (20) can be rewritten

$$\ln u = \ln\left(\frac{d_1kT}{Lh}\right) - \frac{A+BT}{kT} + \frac{\beta\sigma}{kT}, \tag{21}$$

where

$$A = ka,$$

$$B = kb + k\ln\frac{d_1k}{Lh},$$

$$\beta = \gamma kT.$$

The simultaneous validity of (19) and (21) requires that $\tau_b$ be a linear function of stress. In addition, the condition $\tau_b = 0$ for $\sigma = 0$ which should apply at least to annealed metals, yields

$$\tau_b = F(T)\sigma, \tag{22}$$

where $F(T)$ depends only on temperature and material parameters. Equation (21) shows that three factors determine the creep rate, the constants $A$ and $B$ and the temperature function $\beta$. Comparison of (21) with (19) shows that these factors are expressed in terms of physical constants of the material as

$$A = VG_0f^2x^2, \tag{23}$$

$$B = C - Vf^2x^2G_0\alpha, \tag{24}$$

$$\beta = qVxf(1-2F(T)), \tag{25}$$

where the modulus of rigidity $G$ has been taken in the approximate form

$$G \doteq G_0(1-\alpha T), \tag{26}$$

where $G_0$ is the modulus of rigidity at absolute zero.

It seems clear that the structure sensitivity of the creep property, which is particularly important in alloys, will be mainly due to the back

---

[11] P. G. McVetty, Trans. A.S.M.E. **65**, 761 (1943).

[12] S. Dushman, L. W. Dunbar, and H. Huthsteiner, J. App. Phys. **15**, 108 (1944).

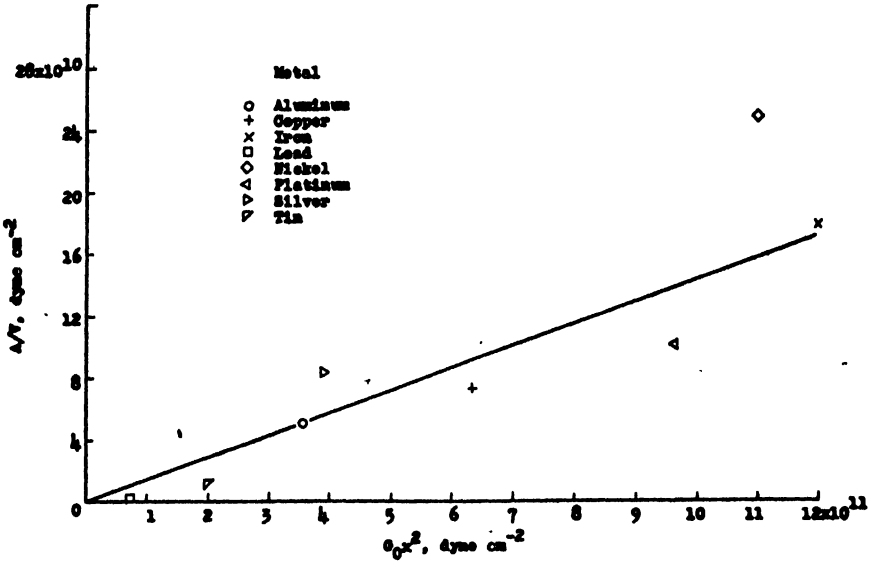


FIG. 3. Comparison of experimental data for factor $A$ with theory.

stress term and will therefore appear in the factor $\beta$.

## COMPUTATIONS AND PRECISION

Creep data were obtained from the literature[13] for a number of polycrystalline "pure" metals that had been annealed. The reproducibility of the creep data from the various sources is not known. The data were plotted up similarly to Dushman and the factors $A$, $B$, and $\beta$ obtained. The factor $d_1/L$ was taken as $10^{-4}$ for all materials.

Generally, the errors in $A$, $B$, and $\beta$ were estimated to be about $\pm 5\times 10^{-13}$ erg, $\pm 8\times 10^{-16}$ erg/°K, and ±30 percent, respectively.

Room temperature values of $G$ were used as approximations to $G_0$.

## RESULTS AND DISCUSSION

### The Factor A

In Fig. 3, $A/V$ is plotted against $G_0x^2$. The line drawn is the best straight line through the origin and generally represents the data within the estimated error of $A$. From Eq. (23) and the above discussion the slope of the line is $f^2$ and should be somewhat less than 0.25. Because the activation energy predicted on the basis of the validity of Hooke's law will be greater than the actual value, it can be expected that the theoretical prediction will be too high. The experimental slope from the straight line of Fig. 3 is 0.14. This agreement is good considering the assumptions made and variety of sources of data used. It gives further verification that the generated dislocation is only one atomic distance in length.

### The Factor B

The order of magnitude of the term $Vf^2x^2G_0\alpha$ of Eq. (24) which is $3\times 10^{-16}$ erg/°K, is within the experimental error of $B$, $\pm 8\times 10^{-16}$ erg/°K; this term can therefore be neglected in consider-

TABLE I. Results for the parameter $B$.
(The values are given in units of $10^{-16}$ erg/°$K$.)

| Metal | $B$ |
|---|---|
| Aluminum | 40 |
| Copper | 45 |
| Iron | 39 |
| Lead | 54 |
| Nickel | 33 |
| Platinum | 42 |
| Silver | 36 |
| Tin | 47 |

---

[13] Aluminum, platinum, and silver: Dushman *et al.*, reference 12; copper: E. A. Davis, J. App. Mech. 10, A101 (1943); H. L. Burghoff and A. I. Blank, Trans. A.I.M.E. 161, 420 (1946); nickel: A. Michel and J. Cournot, Congrès International pour l'Essai des Matérieux (Amsterdam) 1, 397 (1927); iron: A.S.T.M.—A.S.M.E., *Compilation of Available High-Temperature Creep Characteristics of Metals and Alloys* (1938), p. 22; zinc: W. M. Peirce and E. A. Anderson, Trans. A.I.M.E. 83, 560 (1929); tin and lead: L. C. Tyte, Proc. Phys. Soc. 50, 153 (1938); *ibid.* 51, 203 (1939).

ation of the $B$ values. Therefore $B$ becomes equal to $C$, or from (17) to the negative of the entropy of activation $\Delta S_a$. The results obtained from the data are listed in Table I.

A correlation exists between the variation from the best straight line of data presented in Fig. 3 and the variation in $B$ values. The points above the line are all related to metals that have $B$ values less than 40 and those below the line to $B$ values greater than 40. From the method of plotting the data to obtain $A$ and $B$ values, a positive error in $A$ means a negative error in $B$. This correlation therefore indicates that the correct values of $B$ fall about their mean value more closely than the values in the table. In addition, errors in the substitution of $10^{-4}$ for $d_1/L$ and $kT/h$ for $1/t$ will enter into the $B$ factor. The entropy of activation $\Delta S_a$ therefore seems to be very closely the same for all materials considered. This is in agreement with the theory since the factor $p$ should be about the same for all materials.

The large negative value of the entropy of activation $\Delta S_a$ has previously been pointed out by Kauzmann and by Dushman, but satisfactory explanations for the magnitude of this factor could not be given. In terms of the present treatment, the large negative $\Delta S_a$ is accounted for by the restriction of direction of 5 atoms participating in the generation of an activated complex. A value of about $3\times 10^{-3}$ is obtained for $p$ when $C=40\times 10^{-16}$ erg/°K. The possibility that more than 5 atoms must be restricted in direction certainly exists, so that $p$ may be larger than the above value.

### The Factor $\beta$

For a large number of materials the temperature dependence of $\beta$ could be approximated by the function

$$\beta = C_1 \exp C_2 T,$$

where $C_1$ and $C_2$ are positive constants. This function was used to extrapolate to room temperature values.

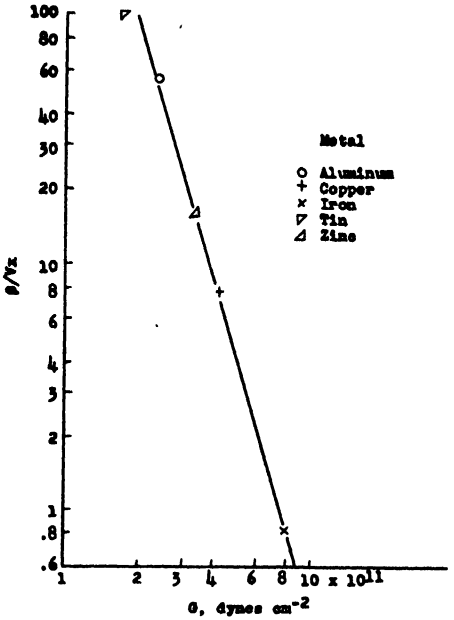


FIG. 4. Room temperature dependence of $\beta/Vx$ on the modulus of rigidity $G$.

The term $F(T)$ in (25) is related to the back stress $\tau_b$ which results from the internal stress field about a dislocation. This stress field is dependent on $G$. Therefore, $\beta$ at any given temperature was expected to be a function of $G$. Figure 4 shows that, for the materials shown, the $\beta$ values extrapolated to room temperature vary as $G^{-n}$ where $n$ is approximately 3.

## CONCLUSIONS

The analysis presented in this report shows that dislocations are generated by the formation of an activated complex configuration in a small region. Limitation of the direction of motion of a group of atoms results in a large negative entropy of activation for the process. When generated, the dislocations are probably just one atom long.

# Application of the Sliding Thermocouple Method to the Determination of Temperatures at the Interface of a Moving Bullet and a Gun Barrel*

E. L. ARMI,† J. L. JOHNSON, R. C. MACHLER, AND N. E. POLSTER
*Leeds and Northrup Company, Philadelphia, Pennsylvania*

(Received May 31, 1946)

Description is given of a set-up for the determination of temperatures at the frictional interface between a moving bullet and the bore in a caliber 0.50 gun barrel. Cathode-ray oscilloscope e.m.f.-time records, which were obtained from a traveling thermocouple junction formed by the bullet-bore interface, are interpreted (by use of a formula due to Jaeger) in terms of instantaneous temperatures. A calorimetrically measured value of the total frictional heat input caused by a bullet being pushed through a barrel section by means of a falling weight, is found to be in good agreement with the value of the heat input calculated from bore surface temperatures which were determined by the sliding thermocouple method.

## INTRODUCTION

DURING the course of an investigation on values of local heat input associated with the firing of individual rounds of ammunition in caliber 0.50 machine gun barrels, it became desirable to obtain an idea of the magnitude of that part of the heat which is caused by friction between bullet and bore, and also an idea of the bore surface temperatures which are reached during this frictional process, before the onset of heat input from the hot powder gases.

A falling weight apparatus with a usable free fall of 65-foot length was modified to drive bullets with fair velocity through a short length of barrel held in a suitable mounting. Making use of the fact that the frictional interface between bullet and bore is also an e.m.f.-generating interface, because of a difference in the thermoelectric potential of the two materials involved (gun steel, gilding metal), effective temperatures at the interface were determined from a measurement of the e.m.f. from this somewhat complex "sliding band thermocouple." Having found the approximate temperature-time curve of a point on the bore, it was easy to calculate the local frictional heat input into the barrel wall. On the other hand, this heat input was also measured directly by a calorimetric method, and thus a check was obtained on the results of the sliding thermocouple method.

Sliding thermocouple measurements were also carried out on caliber 0.50 barrels in actual firings, but the check against calorimetric heat input measurements had to be omitted in this case, because of the impossibility of separating the frictional heat input from that of the powder gases.

## FALLING WEIGHT APPARATUS

Figure 1 gives a view of the falling weight apparatus, with the carriage in raised position. Two taut wire rope guides reached up to a height of 65 feet above a solid concrete foundation. The carriage, sliding on four bearings, was pulled up by a rope and was dropped with the aid of an electromagnetic release mechanism. The weight of the carriage, including the steel bar at its bottom, amounted to 40 lb; occasionally, additional bars were clamped onto the carriage in order to increase the kinetic energy of the system. At the end of its fall the carriage hit a piston which in turn drove the bullet through a short length of barrel, held in a heavy steel base. Excess kinetic energy was taken up by lead stops of suitable size.

Electrical connection to the bullet was made by means of a shielded flexible cable soldered to the tip of the bullet; the grounded shield of the cable was connected to the barrel section. E.m.f.'s generated at the frictional interface were amplified in a Ballantine Decade Amplifier (Model 220) before being made visible on the screen of a DuMont cathode-ray oscillograph

---

* The work reported in this article was undertaken as part of contract OEMsr-536 between the Leeds and Northrup Company and the Office of Scientific Research and Development, under the supervision of Division One of the National Defense Research Committee.

† Now Director of Technical Department, Mica Insulator Company, Schenectady, New York.

(Type 175A). The latter was operated with single sweeps which were obtained from a DuMont low-frequency linear time base Generator (Type 215) and which were triggered by the closing of a contact shortly before the carriage hit the piston driving the bullet. Figure 1 shows the double spike on the carriage, which closed the sweep initiating circuit of the time base generator by connecting the two mercury pools in the reservoirs visible below the spikes.

Details of the construction of the barrel section holder can be seen in Fig. 2. (The thermocouple shown on the outside of the barrel section was used for the calorimetric measurements discussed under the heading "Theory of Calorimetric Measurements in Barrel Sections.") In most tests the barrel sections had special 0.500″/0.520″ rifling (bore diameter 0.500″ across lands, 0.520″ across grooves) instead of the regular 0.500″/0.510″ rifling of caliber 0.50 guns, in order to avoid contact of the bullets with the bottom of the grooves of the rifling; the generation of thermal e.m.f. was thereby restricted to the area of the top and (part of) the sides of the rifling bands ("lands"), and the interpretation of the records obtained became somewhat simpler. All barrel sections had at their entrance a forcing cone (origin of rifling) of standard taper.

## INTERPRETATION OF CRO TRACES OF THE INTERFACE THERMAL E.M.F.

Interpretation of the experimental data was greatly helped by a paper of J. C. Jaeger, on "Moving Sources of Heat and the Temperatures at Sliding Contacts,"[1] in which temperatures in a semi-infinite conductor are worked out for the motion of a band source of heat over the conductor, and this analysis is applied to the case where a slider moves with constant friction over a metallic surface. Jaeger shows that, for high

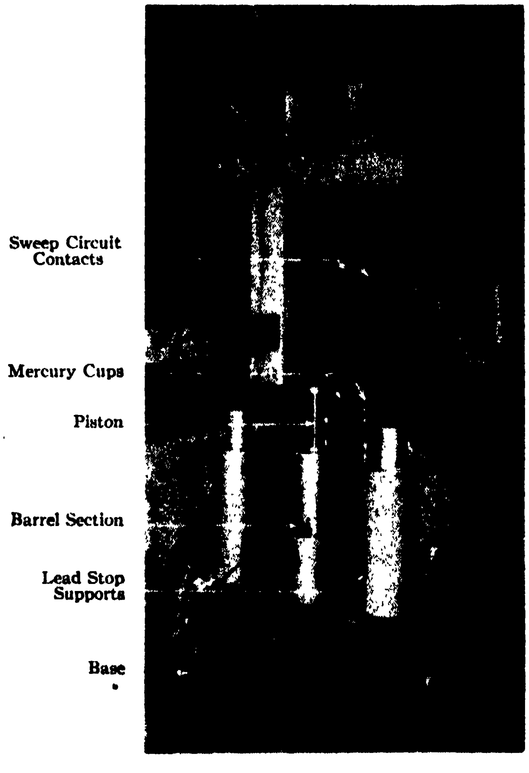


FIG. 1. Falling weight apparatus (carriage in raised position).

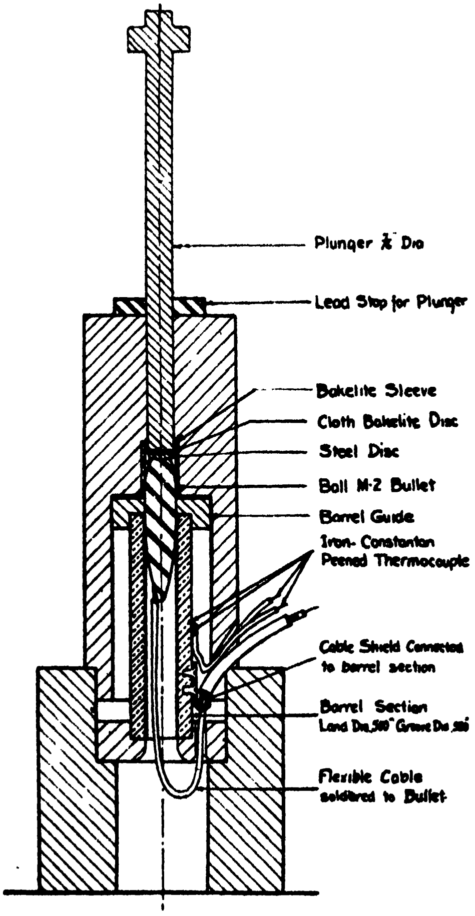

Plunger ⅞" Dia
Lead Stop for Plunger
Bakelite Sleeve
Cloth Bakelite Disc
Steel Disc
Ball M-2 Bullet
Barrel Guide
Iron-Constantan Peened Thermocouple
Cable Shield Connected to barrel section
Barrel Section Land Dia .500" Groove Dia .520
Flexible Cable soldered to Bullet

FIG. 2. Details of mounting of barrel section.

---

[1] J. C. Jaeger, Proc. Roy. Soc., N. S. W. 76, 203–224 (1943).

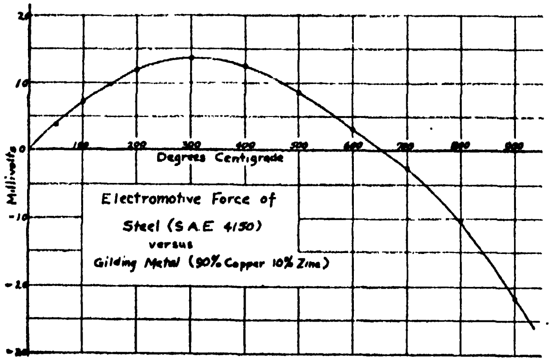


FIG. 3. E.m.f. calibration curve of gun steel *vs.* gilding metal.

sliding speeds and after a steady state is reached, the instantaneous temperature distribution over the surface of the metal in contact with the slider can be approximated by[2]

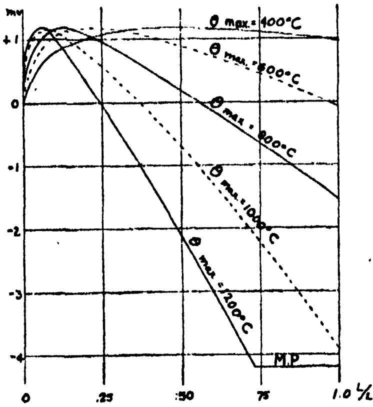
$$\theta/\theta_{\mathrm{max}} = (L/L_0)^{\frac{1}{2}}, \qquad (1)$$

where $\theta$ is the temperature rise on the surface of the metal, $L$ the distance back from the front edge of the "band source," and $L_0$ the total width of the "band source." It seems permissible, for a first approach to the solution of the problem, to assume that the conditions of the theory are satisfied to an acceptable degree in the case of a bullet sliding over the (spirally rifled) bore surface, whereby the frictional contact area corresponds to the heat-generating band of the mathematical analysis.

Matters become complicated by the behavior of the e.m.f.-temperature curve of the particular combination of metals in the sliding thermocouple (gun steel *vs.* gilding metal). Figure 3 shows a calibration curve taken on a sample thermocouple, the two wires of which were cut from a bullet jacket and a gun barrel, respectively. This curve, which has its maximum at about 300°C, is parabolic in its general shape; all e.m.f. readings are, therefore, essentially ambiguous since their correlated temperatures can lie either on the ascending or the descending branch of the calibration curve. It should also be kept in mind that the e.m.f. recorded is essentially an arithmetical average of the e.m.f.'s generated over the whole frictional interface area at the instant concerned; but, because of the nonlinearity of the e.m.f.-temperature curve, the temperature corresponding to the average of the e.m.f.'s is not at all identical with the average of the temperatures which cause the generation of these e.m.f.'s.

Figure 4, based on Eq. (1) and Fig. 3, gives curves of the e.m.f.'s which are generated locally along the frictional interface. Of course, these e.m.f.'s are never measured as such, since only the average over all locally generated e.m.f.'s is recorded on the CRO screen. Figure 5, above the designation "full contact," shows this average for several values of $\theta_{\mathrm{max}}$. However, it should be remembered that the "local" e.m.f.'s are actually averages too, namely over-all points which make contact with the bore on a particular circumference of the bullet. (It is therefore, in general, quite possible to find points which have temperatures in excess of the effective "maximum temperature.") Since the temperature in the frictional process is unlikely to exceed the melting point of the material of the bullet jacket (gilding metal), the curves for a fictitious $\theta_{\mathrm{max}} = 1200°$ have to stop at the e.m.f. corresponding to the temperature of that melting point (about 1050°C).

As long as the bullet is fully in contact with the bore surface, averaging of the e.m.f.'s takes place over the whole of the curves of Fig. 4. However, when the bullet is entering or leaving the barrel section, the average is taken over only that part of the curve which corresponds to actual contact with the bore. Figure 5 gives e.m.f.



LOCAL EMF ON MOVING INTERFACE THERMOCOUPLE
(COLD JUNCTION 25° C)

FIG. 4.

---

[2] J. C. Jaeger, reference 1, Eq. (14).

curves which were obtained by carrying out the averaging as indicated, for the bullet entering, traversing, and leaving the barrel section, with $X/L_0$ denoting the ratio between instantaneous length of contact and length of contact for a fully entered bullet. The relative length of the center part of the curve, designated "full contact," varies, of course, with the length of the barrel section used.

## COMPARISON BETWEEN THEORETICAL AND OBSERVED FORMS OF OSCILLOGRAM RECORDS

Figure 6 shows three frictional interface e.m.f. traces, together with theoretical curves for $\theta_{\max} = 825°C$ and 910°C, respectively, for a $3\frac{5}{8}$ in. long barrel section. Disregarding irregularities of detail, these traces are seen to conform well with the theoretical curves; in fact, it is almost surprising that in spite of the simplifications made for the computation of the theoretical curves, such satisfactory agreement with experimental work was achieved. The maximum temperatures ($\theta_{\max}$) encountered on these and other similar records are found to be between 800°C and 950°C.

Speculating on the causes for the irregularities in the traces, it can be said that the vibrations seen on most traces are probably caused by a longitudinal oscillation of the bullet; careful degreasing of bullet and barrel had the effect of accentuating this vibration as well as of raising the interface temperature. A local increase in frictional pressure causes additional heat development and, consequently, a rise in the interface temperature; it is therefore probable that some of the deviations from theoretical curves are due to variations of pressure. The passage of the crimped portion of the bullet jacket might be expected to be noticeable in the traces, and, indeed, the secondary peak in most exit traces may be attributed to it. Finally, since interface temperatures go up with an increase in the velocity of the bullet, the traces for 64-foot drops are seen to correspond to slightly higher temperatures than those for 32-foot drops. In the

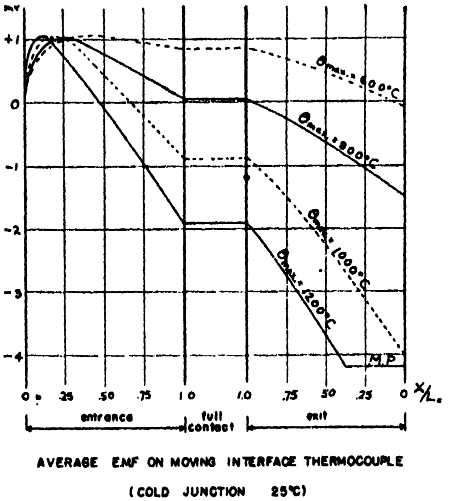


FIG. 5.

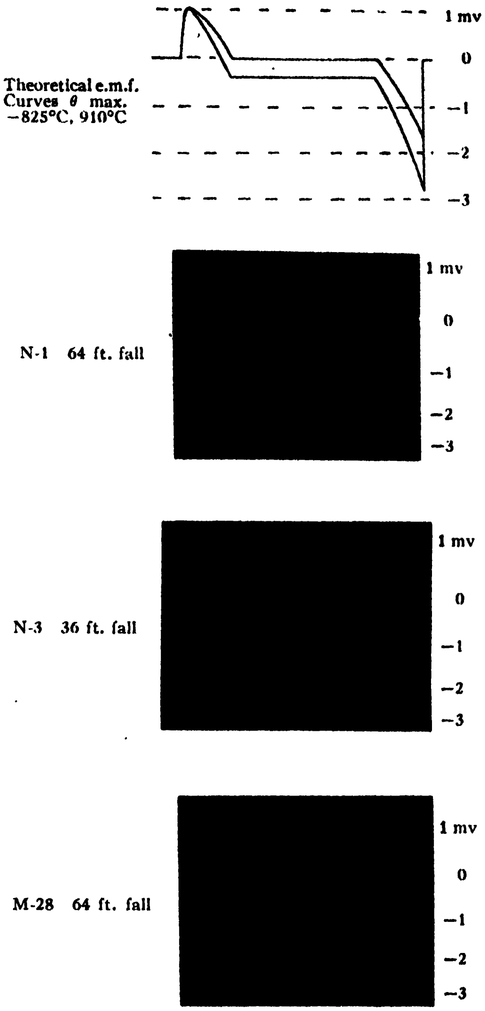


FIG. 6. Interface thermal e.m.f. traces taken with falling weight apparatus. Barrel section No. 4 (0.500"/0.520")—40-lb. carriage. Timing dot intervals 1/5000 sec. Oiled M2 bullets.

first approximation, the interface temperature rises with the square root of the velocity,[3] but, at higher temperatures the coefficient of friction decreases, and the interface temperature eventually levels off. In the case of drops with low initial kinetic energy, where the bullet velocity decreases appreciably during the travel through the barrel section, the e.m.f. traces observed were found to be distorted in accordance with the expected decline in the interface temperatures.

Figure 7 shows four records which were taken on a 2-in. long barrel section. These records are quite similar to those seen in Fig. 6, but, since they were taken on a much shorter barrel section, the flat central portion of the traces is correspondingly shortened. While the general agreement among traces which were taken under similar conditions is quite satisfactory, it is not surprising to find a number of minor variations between individual traces, due to small changes in local conditions between individual rounds.

## INTERFACE THERMAL E.M.F. TRACES FROM FIRED BULLETS

After the work with the falling weight apparatus had given encouraging results, an attempt was made to obtain records of the frictional interface e.m.f. for actually fired bullets. The problem of keeping an insulated wire attached to the tip of the moving bullet was considerably simplified when the barrels were cut down to only 6 in. of rifled length. Experiments undertaken with these short barrels gave quite satisfactory results.

Since it was found desirable to correlate the recorded instantaneous thermal e.m.f. with the position of the bullet, a simple apparatus was devised to record the time when the bullet passed through certain points. A number of hardened steel wires were placed at fixed distances in an insulated holder over the muzzle end of the shortened barrel. A disk mounted on a "harpoon" which extended from the front end of the bullet, touched and broke these steel wires in succession, as the bullet moved forward, and thereby closed and opened a circuit which applied a fixed voltage to the CRO. A leaf-spring contact, mounted at the rear end of the barrel receiver was automatically closed by the motion of the firing mechanism and thereby made to initiate a single sweep on the CRO; the closing and opening of the circuit through the bullet harpoon and the

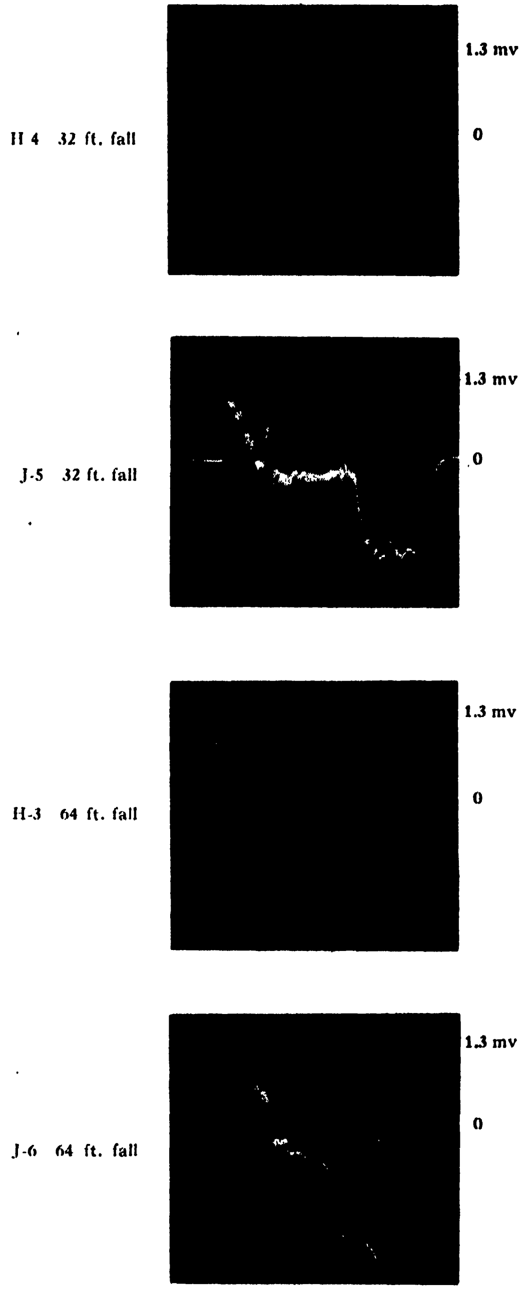


FIG. 7. Interface thermal e.m.f. traces taken with falling weight apparatus. Barrel section No. 3 (0.500″/0.520″)—oiled M2 bullets. Timing dot intervals 1/5000 sec.—40-lb carriage.

---

[3] J. C. Jaeger, reference 1, Eq. (33).

steel wires, during the sweep, was used to obtain position-time records.

The electrical set-up for taking frictional interface e.m.f. traces during the firing was the same as that for the falling weight apparatus, with the above-mentioned contact on the barrel receiver taking the place of the spikes on the carriage and the twin mercury pools as switching device for the initiation of a single sweep on the CRO.

Traces obtained with a fired bullet are essentially similar to those from the falling weight apparatus, however, because of the much higher velocity of a fired bullet, fusing is found to occur over a considerable portion of the frictional area; the simple parabolic temperature distribution over the length of the area of contact, which is valid with a constant coefficient of friction, does therefore not hold for this case. Furthermore, since the CRO traces give records of e.m.f. *vs.* time rather than *vs.* travel, the traces obtained have to be redrawn with the use of position-time records, if records of e.m.f. *vs.* travel are desired.

## THEORY OF CALORIMETRIC MEASUREMENTS IN BARREL SECTIONS

In addition to the recording of interface thermal e.m.f.'s, the experimental set-up was well suited and easily adapted for calorimetric measurements of the amount of frictional heat transferred to the barrel wall. It appeared of interest to carry out such measurements and to compare the results obtained with heat input values which were calculated from bullet-bore interface temperature determinations.

A heat input $H$ per unit area of bore wall, in a given unit length of barrel (of annular cross section), causes a temperature rise $\theta_e$ throughout that length amounting to, after temperature equilibrium has been reached,

$$\theta_e = \frac{H}{bc\rho}\frac{1}{1+b/d}, \tag{2}$$

where $b$ is the thickness of the barrel wall in the section, $d$ the bore diameter, $c$ the specific heat of the barrel material, and $\rho$ its density.

The time required to reach thermal equilibrium throughout the barrel section varies approximately with the square of the thickness of the barrel wall,[4] and, *e.g.*, for a wall thickness of $\frac{1}{4}$ in., it is of the order of 2 sec.

While the reading of the "barrel calorimeter" is not affected by a continuous variation in the heat input over the length of the barrel, a discontinuity in the amount of the heat input, or a change in the cross section of the barrel, or any other "end effect" does influence the temperature in the cross section in which the measurement is taken. Thus, a temperature difference $\delta_e$, which is maintained at a distance $x$, causes a temperature difference $\delta$ in a cross section of thermal diffusivity $a^2$, after a time $t$, according to[5]

$$\delta/\delta_e = 1 - \mathrm{erf}(x/2at^{\frac{1}{2}}). \tag{3}$$

Therefore, although the calorimetric measurement cannot be made before radial temperature equilibrium has been attained, it is essential that the reading or recording be completed before any temperature difference due to axial flow becomes appreciable. *E.g.*, a value of $\delta/\delta_e = 0.0025$ is reached in steel after about 25 sec. with $x = 3$ in.; with $x = 1$ in., the same value is reached after slightly less than 3 sec.

Depending on the time required to carry out the temperature measurement (whether done manually or by some automatic recording equipment) the distance $x$ has to have a definite minimum value. *E.g.*, for heat input determinations in caliber 0.50 firings, the distance of a thermocouple from the end of the "calorimetric section" of reduced outside diameter could be kept as short as 1 in., when measurements were made with a Leeds and Northrup Speedomax A (electronic potentiometer recorder); this distance allowed a time of about 1 sec. between the establishment of radial temperature equilibrium and the beginning of a perceptible temperature decay in the section, caused by heat flow to the ends of the "calorimetric section."

The amount of heat received by any point on

---

[4] $$\theta/\theta_e = 1 + \sum_{n=1}^{\infty} (-1)^n \exp(-n^2\pi^2a^2t/b^2).$$

See H. S. Carslaw, *Mathematical Theory of the Conduction of Heat in Solids* (The Macmillan Company, London, 1921), second edition, Chap. IX.

[5] H. S. Carslaw, reference 4, Chap. III.

the bore during passage of a bullet is given by[6]

$$H=\left(\frac{kc\rho}{\pi}\right)^{\frac{1}{2}}\int_0^{\tau}\frac{\theta(t)}{(\tau-t)^{\frac{1}{2}}}dt, \qquad (4)$$

where $H$ is the heat input per unit area of bore, $k$ the thermal conductivity, $c$ the specific heat, $\rho$ the density of the barrel material, $\theta$ the temperature rise of the bore as a function of time $t$, and $\tau$ the total time of contact of the point concerned.

Assuming that it is permissible to approximate the temperature distribution over the contact area of the bullet with the bore by Eq. (1), which is valid for a "slider" moving under steady-state conditions, the temperature of a fixed point on the bore, during passage of the bullet, is given by:

$$\theta(t)=\theta_{\max}(t/\tau)^{\frac{1}{2}}. \qquad (5)$$

Substituting (5) into (4) and integrating, one obtains finally:

$$\theta_{\max}=\frac{2}{\pi^{\frac{1}{2}}}\frac{H}{\tau^{\frac{1}{2}}(kc\rho)^{\frac{1}{2}}}. \qquad (6)$$

## RESULTS OF CALORIMETRIC MEASUREMENTS

For heat input measurements in the falling weight apparatus, the "calorimetric section" had a length of 3⅝ in., with the couple peened in its center; no thermal insulation at the ends of the section was required. A special Leeds and Northrup mirror galvanometer recorder was used as recording equipment.

The frictional heat input values obtained by the "calorimetric section" method scatter rather widely, averaging 2.85 cal./cm² and 3.65 cal./cm² for drops from 64 feet and 32 feet height, respectively, when referred to the whole bore surface; when only the area of actual contact (assumed to consist of the top and about one-half of the sides of the lands) is taken into account, these values rise to 8.7 cal./cm² and 11.0 cal./cm², respectively, in a 0.500″/0.520″ rifled barrel section, at about 1¾ in. beyond the origin of rifling.

When these heat input values are substituted in Eq. (6), with $(kc\rho)^{\frac{1}{2}}=0.31$ cal./°C cm² sec.$^{\frac{1}{2}}$ (for steel) and a length of contact $L_0=2.0$ cm, values of $\theta_{\max}$ between 950°C and 1000°C are obtained for either case (with bullet velocities of about 60 ft./sec. and 40 ft./sec., respectively). These temperatures are somewhat higher than those calculated from sliding thermocouple records (800°C to 950°C); however, in view of the simplifying assumptions made for the calculations and of the various uncertainties involved, the agreement can be considered quite satisfactory. Furthermore, the errors in the temperatures calculated are in the direction to be expected; while Eq. (4) is based on the assumption of linear heat flow from the bore surface into the barrel, heat actually is generated over only a fraction of the bore surface (top and about one-half of sides of lands) in barrels used for this work, and it spreads from there two-dimensionally, with the effect that a lower temperature than with strictly linear flow is required at the contact area in order to produce a given heat input; this effect increases rapidly with duration of the time of contact, and is therefore more pronounced for 32 foot drops than for 64 foot drops.

Concluding, it may be said that frictional temperature determinations by means of the sliding thermocouple method gave satisfactory results in an application where no other known methods of temperature measurement could be used.

---

[6] See H. S. Carslaw, reference 4, p. 47, and note that $H=\int_0^{\infty}c\rho\theta_x dx$, where $\theta_x$ denotes the temperature at any point $x$.

# Carbide Structures in Carburized Thoriated-Tungsten Filaments

C. W. HORSTING
*Tube Department, Radio Corporation of America, Harrison, New Jersey*

(Received April 16, 1946)

A wide variety of carbide structures is described, as found in the surface layer of carburized thoriated-tungsten filaments. Their origin is traced back to carburizing conditions and subsequent processing during tube making. A frequently occurring laminated carbide structure is found to contain less carbon than $W_2C$. The thyratron relay method of carburizing control is critically reviewed. Abnormal filament current in manufactured tubes is explained as due to surface conditions which cause changes in thermal emissivity.

THE application of carburized tungsten-filament cathodes is a well-known practice in radio tube manufacture, especially in the power tube field where high electron emission is required at high voltages. By the use of this method, longer filament life and higher operating temperatures are possible because, it is claimed, the emission-causing thorium layer at the filament surface adheres more firmly to the tungsten-carbide surface than to the pure tungsten surface.[1]

During the production of power tubes, the tungsten filaments are partially converted into tungsten carbide by heating them electrically in a hydrocarbon atmosphere to a temperature of about 2200°C. The hydrocarbon is decomposed at the hot filament surface to produce free carbon which diffuses into the tungsten, forming tungsten carbide. The degree of carburization is limited by the fact that the mechanical strength of the filament decreases rapidly with increasing carbon penetration. In practice, therefore, the carburizing reaction has to be stopped accurately at a predetermined penetration.

For this purpose, and chiefly because of the high speed of the carburizing reaction, a thyratron-controlled relay is often employed to cut off the heating current and, thus, stop the reaction after the desired amount of carbide has been formed. The fact that tungsten carbide has a higher electrical resistance than tungsten is utilized to determine the action of the relay. Since carburization is allowed to take place at a constant heating current, a rising voltage appears across the filament terminals as the resistance of the filament increases. The thyratron-operated relay is adjusted to function at a desired point on the rising voltage curve. Ideally, this point corresponds to a constant degree of carburization.

However, under certain circumstances hitherto unknown, filaments were produced which showed abnormal properties. The main difficulty encountered was too large a spread of the cold-resistance and the hot-resistance values of the finished tubes and inconsistent correlation between cold and hot resistance values.

In order to determine the reasons for these variations, the following investigation was carried out. At the start of the investigation, the main questions could be briefly stated as follows: Why does the theoretically well-founded method of thyratron control fail to produce tubes with constant filament properties? What are the circumstances under which the thyratron control fails? Is the resistance-control method unsuitable? What secondary phenomena may be causing the discrepancies?

In order that answers to these questions could be obtained, filaments having abnormal characteristics were collected and analyzed metallographically as to their microstructure. A wide variety of structures was found. Partial cross sections of two of these filaments are reproduced in Figs. 1 and 2. Figure 1 shows a filament which was found to be normal, whereas Fig. 2 shows a filament with normal cold resistance, but with low hot resistance, and a resultant filament current 6 to 7 percent above the value normally occurring at the specified operating voltage.

It has been well-known that different structures and forms of tungsten carbide can be produced under different circumstances. W. P.

---

[1] L. R. Koller, *The Physics of Electron Tubes* (McGraw-Hill Book Company, Inc., New York, 1937).

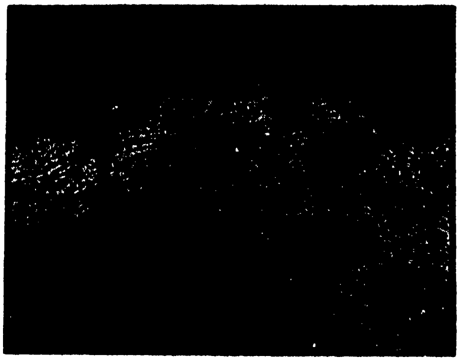


FIG. 1. Normal carbide structure. Magnification 250×.

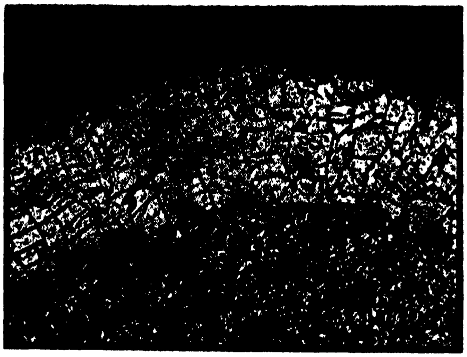


FIG. 2. Abnormal carbide structure. Magnification 250×.

Sykes,[2] K. Becker,[3,4] B. T. Barnes,[5] M. R. Andrews,[6] and others have reported on a variety of carbide structures, but the relation between the different carbide formations present in carburized tungsten filaments and their origination from the carburizing process had hitherto not been established. Consequently, efforts were made to reproduce under laboratory conditions the formations found in filaments and, thus, to determine the factors that produced them.

For that purpose the apparatus shown in Fig. 3 was constructed. With this apparatus, a U-shaped filament can be electrically heated in a hydrogen atmosphere to which can be added a controlled amount of hydrocarbon vapor by means of a specially devised carburetor. The current through the filament is controlled by means of a Variac auto-transformer. The carburizing time is regulated by a time switch which operates a relay switch in the main circuit. Such filaments carburized with this apparatus gave accurate reproduction of carbide formation and of electrical resistance.

One of the first filaments produced with this apparatus had a carbide structure as shown in Fig. 4. About half of the carbide shell consists of a laminated carbide formation similar to the one found in the normal filament. The other half is of a massive structure. By changing the hydrocarbon content of the hydrogen gas, it was possible to change the proportions of the two formations in the carbide shell at will. By lowering the hydrocarbon content, more of the laminated structure appeared and, simultaneously, more time was required for a certain depth of penetration of the carbide, whereas increasing the hydrocarbon content had the opposite effect. Figures 5a, b, and c show three different samples which were carburized to the same depth but with increasing hydrocarbon concentrations.

Because it is not possible to reproduce photomicrographs of all the cases found, a graphical reproduction of the whole range of structures obtained with a wide variation of hydrocarbon concentrations is given in Fig. 6. As a key to the interpretation of Fig. 6, a graphic picture corresponding with the photograph shown in Fig. 4 is given in Fig. 7. It will be explained later on how the concentration-penetration diagrams were arrived at. In Fig. 6, the diagrams of carbon

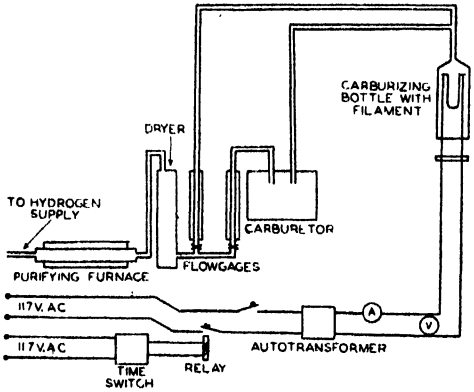


FIG. 3. Carburizing apparatus.

---

[2] W. P. Sykes, *A Study of the Tungsten-Carbon System.* Paper presented at the 12th annual convention of the Am. Soc. for Steel Treating, Chicago, 1930.

[3] K. Becker, "Ueber das system Wolfram-Kohlenstoff," Zeits. f. Metallkunde **20**, 437–41 (1928).

[4] K. Becker, "Die Konstitution der Wolfram-Karbide," Zeits. f. Electrochemie **34**, 640–642 (1928).

[5] B. T. Barnes, "Properties of carbonized tungsten," J. Phys. Chem. **33**, 688–691 (1929).

[6] M. R. Andrews, J. Phys. Chem. **27**, 275 (1923).

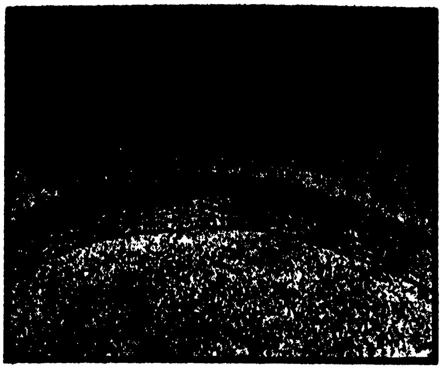


FIG. 4. Carbide shell containing solid as well as laminated carbide structure. Magnification 250×.

concentration *versus* penetration correspond to the structures shown at the left.

Figure 6a shows the effect produced by carburizing a filament in an atmosphere rich in hydrocarbon. A carbon deposit covers the surface, and two massive carbide phases are separately visible. The inner one is almost pure $W_2C$, whereas the outer one is WC. The concentration-penetration diagram for this sample is shown on the right.

Figure 6b is the same but without the carbon deposit.

Lowering the hydrocarbon content somewhat more prevents the formation of the second solid phase, the result of which is shown in Fig. 6c.

In Fig. 6d, we see the appearance of the laminated phase which was found in the normal filament; Figs. 6e–6h show the effects of reducing the hydrocarbon content further. The massive structure decreases until only widely spaced laminated material is formed at the very lowest hydrocarbon concentration.

The fact that the laminated structure is connected with a low hydrocarbon concentration, and the massive structure with a higher one, suggests that the laminated material contains less carbon than the massive $W_2C$ phase.

In addition, we know that where only laminated material is formed, the layers become wider as the hydrocarbon content of the carburizing atmosphere decreases.

Combination of these two facts leads to the conclusion that the change in character which we observed in the laminated phase in Fig. 4 actually shows us a concentration gradient of decreasing carbon content from the massive phase toward

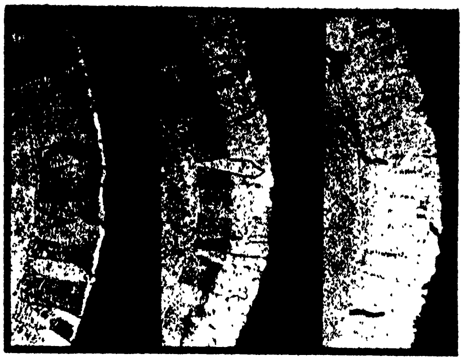


FIG. 5. a, b, and c. Carbide shells produced by carburization to the same depth of penetration, but with increasing hydrocarbon content of the hydrogen atmosphere. Magnification 175×.

the core. The gradual transition between the massive $W_2C$ phase and the laminated phase

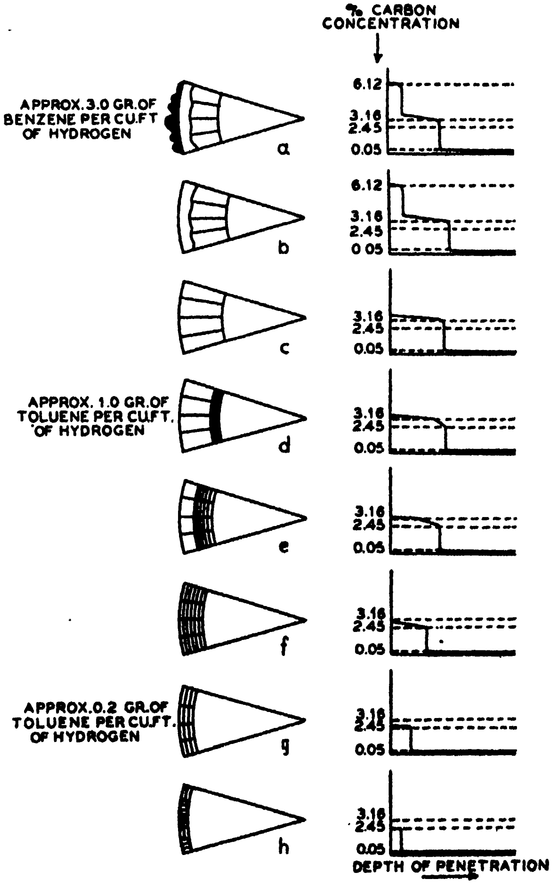

% CARBON CONCENTRATION
APPROX. 3.0 GR. OF BENZENE PER CU. FT OF HYDROGEN
APPROX. 1.0 GR. OF TOLUENE PER CU. FT. OF HYDROGEN
APPROX. 0.2 GR. OF TOLUENE PER CU. FT. OF HYDROGEN
a
b
c
d
e
f
g
h
6.12
3.16
2.45
0.05
DEPTH OF PENETRATION

FIG. 6. Carbide formations produced by carburization with constant carburizing time and constant starting temperature, but with decreasing hydrocarbon content of the hydrogen atmosphere.

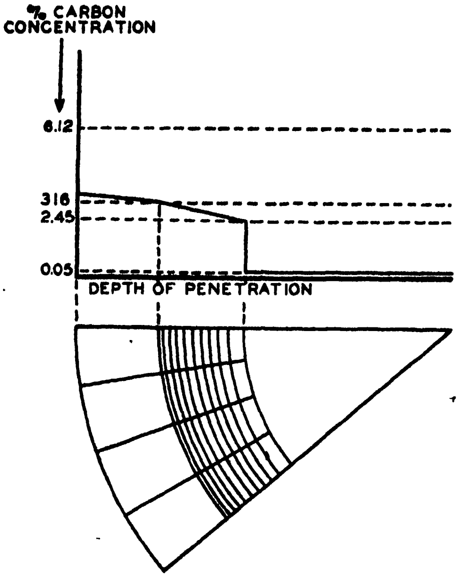


FIG. 7. Key to the interpretation of Figs. 6 and 11.

suggests that the carbon deposit at this side of the gradient starts at 3.16 percent. The sharp line of demarcation between the laminated phase and the core suggests that the gradient does not end in a zero carbon concentration, but abruptly drops from a certain value to zero or some very small value.

In order to test the idea of a gradient, it was reasoned that a gradient of reverse character should be produced if carbon were removed from a previously carburized filament. This was done by heating such a filament in moist hydrogen. Figure 8b shows the results obtained. Figure 8a shows the original structure. The partial removal of carbon from the filament produced a laminated layer near the surface of the filament. It is clearly visible in the cross section that the laminations are widest near the surface where even a small decarburized layer is visible. The gradient on the inside of the remaining $W_2C$ section is, of course, of decreasing carbon content toward the core. The opposite character of the inward and outward gradients is clearly visible in Fig. 8b.

The explanation of the concentration-penetration diagrams in Fig. 6 is based on the following considerations and experiments.

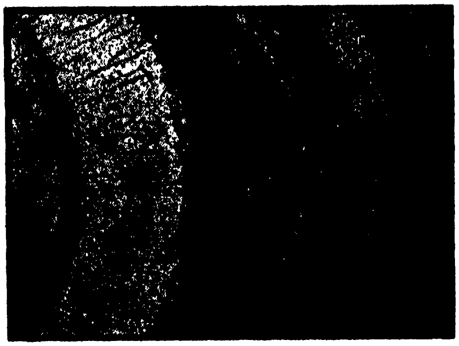
If we briefly review the theory of diffusion in solids, it does not seem to be difficult to explain the structures which are reproduced in Fig. 6.

If diffusion occurs in a binary system, and the components show mutual solubility in all proportions, one can generally expect a smooth concentration-penetration curve showing a gradual concentration decrease from the surface inward and ending in a zero concentration of the penetrating element at some distance from the surface.

However, in case limited solubility occurs, an area of co-existent phases separates the solubility fields in the phase diagram. If diffusion takes place in such a system, a sharp break in the concentration-penetration curve can be expected. The concentration drop at this break corresponds to the difference in solute content of the co-existent phases at the temperature involved, while the absolute concentrations on either side of the break are equal to the solute contents of the two co-existing phases.

In the somewhat more complicated case where compounds are formed between the components of the binary system, several areas of co-existence may occur, which are separated by more or less wide areas of solid solubility. Corresponding to each area of co-existence, a sudden break in the concentration-penetration curve must be expected. Each area of solid solubility corresponds to a gradient of gradually changing concentration.

Let us now consult the available data on the system tungsten-carbon and see what has to be expected in view of the above (see Fig. 9). Two compounds occur between tungsten and carbon, $W_2C$ and WC, the first containing 3.16 percent carbon, the other 6.12 percent carbon. Starting



a b

FIG. 8. Carbide shells. a, before, b, after, partial decarburization in wet hydrogen. Magnification 175×.

from the tungsten side of the diagram and only considering solids, we find successively:

1. A very narrow field of solid solubility of carbon in tungsten (W. P. Sykes reports a maximum solubility of 0.05 percent carbon in tungsten at 2400°C).
2. A co-existent phase area.
3. A narrow field of solid solubility around the stoichiometric $W_2C$ composition.
4. A co-existent phase area.
5. The compound WC with no solid solubility field.
6. An area of co-existence between WC and graphitic carbon.

If we apply the afore-mentioned theory to these data, we can expect the shape of the concentration-penetration curve for carbon diffusing into tungsten to be as indicated in Fig. 10. It is obvious, of course, that the complete sequence of concentration changes from the surface inward can occur only if sufficient carbon is available at the surface, and that the relative amounts of the different formations will depend upon the diffusion velocities within those formations and the degree of equilibrium reached during the diffusion reaction.

The concentration gradient within the solid solution range is, for simplicity's sake, indicated as a straight line. It will generally be a curved line which may be convex or concave to a larger or lesser degree depending on carburizing conditions. With this diagram as a basis, it is comparatively simple to construct the concentration-penetration diagrams, which are given in Fig. 6. If, on the basis of the above analysis, we reconsider the laminated gradient, it becomes clear that this material must be identified with the solid solution range of tungsten in $W_2C$. In order to explain its laminated character we have to invoke Sykes' claim that the solubility of tungsten in $W_2C$ decreases with temperature and that the tungsten-rich phase precipitates out on cooling from the diffusion temperature. Heavier layers are formed where more tungsten was in solution, while the wider spacing of the layers occurring simultaneously, may be attributed to the higher precipitation temperature. That no precipitation seems to occur where excess carbon is dissolved in $W_2C$ would indicate that the solubility of carbon in $W_2C$ does not decrease with temperature or that it is small and shows no pronounced effect on precipitation.

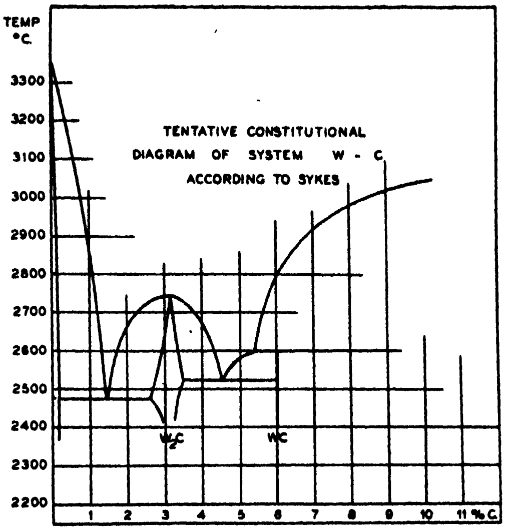


FIG. 9. Tentative constitutional diagram of tungsten carbon (used with permission).

The existence of the carbon concentration gradient, not only in the laminated region, but also in the massive carbide, was made plausible by a series of micro hardness tests. Successive tests, starting from the center of the wire outward toward the surface, were made with a Tukon hardness tester on a sample like the one shown in Fig. 4 containing both the massive and the laminated material. The core was found to be of uniform hardness. A sudden sharp increase in hardness occurred at the demarcation line between core and carbide shell followed by increasing hardness throughout the laminated zone. A maximum hardness was found just inside the massive zone, while the hardness again decreased from there to the surface. The increase in hardness throughout the laminated part of the shell confirms the gradient in this part; the relative amount of $W_2C$ increases. But the hardness distribution in the massive material also indicates an increasing carbon content toward the surface, for it is known that $W_2C$ attains its maximum hardness at the stoichiometric composition. Excess carbon, however, again lowers the hardness.

Finally, a direct carbon analysis of the laminated material was made in order to establish definitely its carbon content. At the same time

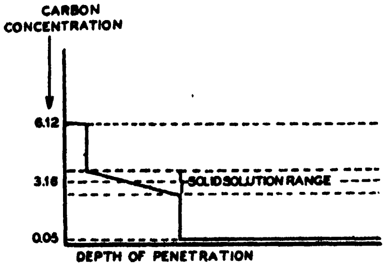


FIG. 10. Theoretical concentration-penetration curve for the diffusion of carbon into tungsten.

an attempt was made to determine the solubility limit of tungsten in $W_2C$ or, what amounts to the same thing, the carbon concentration at the carbide-side of the sharp interface between the carbide shell and the core. If carburization is carried out at the lowest hydrocarbon concentration which will allow formation of laminated material, and the temperature is sufficiently high to guarantee a diffusion reaction which is essentially "carbon-limited," it is safe to assume that equilibrium conditions are closely approached and that whatever gradient is formed is so small as to be negligible. The concentration-penetration diagram of this laminated material is indicated in Fig. 6h, and the whole carbide shell is of a widely spaced laminated character which does not show any evidence of a gradient. This material was produced and analyzed.

The carbon analyses were carried out with the aid of a high vacuum carbon analysis system which was built on the principle first described by Yensen,[7] later improved by Ziegler,[8] and redesigned by Wooten and Guldner[9] and others. The process, briefly described, utilizes a sample weighing about 10 milligrams which is heated in a vacuum to 1000°C and then burned in an excess of purified oxygen. After the $CO_2$ formed is frozen out in a liquid-air trap, the excess oxygen is removed by means of a vacuum pump. The $CO_2$ is then permitted to expand into a known volume. The $CO_2$ pressure at room temperature then furnishes the data required for a calculation of the carbon content of the sample under analysis.

At the same time, cross sections were made of the immediately neighboring part of the filament; the relative areas of the carbide shell and the tungsten core were measured by the use of a planimeter on the tracing of the enlarged cross-section image. This information made it possible to calculate the carbon content in the shell.

Results are given in Table I.

From the results obtained it can be seen that the carbon content is lower than 3.16 percent, as expected and, furthermore, that the equilibrium condition mentioned above was closely approached. This is borne out by the fact that the values obtained are fairly constant, even though the samples were carburized to different depths. We have assumed the average value of 2.45 percent carbon to be the one corresponding to the solubility limit of tungsten in $W_2C$, and the diagrams in Fig. 6 are drawn accordingly.

If we compare Sykes' tentative curves (Fig.9) with the result of our experiment, it seems that the value of 2.45 percent is somewhat low. However, the diffusion method is one of the most sensitive ones for the investigation of solid solubility limits, and Sykes' constitution diagram does not claim great accuracy in this region.

During the study of the decarburization process, it was found that the conditions shown in Fig. 8b could be produced only by high speed decarburization, that is, in wet hydrogen where oxygen is present to react with the carbon and remove it quickly from the filament surface as CO or $CO_2$ gas. In dry hydrogen, decarburization is much slower, and the massive $W_2C$ layer almost entirely disappears as a result of the carbon penetrating inward before a visible gradient toward the outside is established. A general survey of the stages of decarburization reactions is given in Fig. 11.

Decarburization in high vacuum is still slower than in pure hydrogen and needs a higher temperature. In high vacuum only a very limited amount of gas is present for the carbon to react with and simple evaporation may become necessary. Furthermore, it was found that the diffusion velocity of carbon in tungsten is considerably lower in a vacuum than in the presence of hydrogen. However, the same pattern of de-

---

[7] T. D. Yensen, "Carbon in iron," Trans. Am. Electrochem. Soc. 37, 227–245 (1920).

[8] Nicholas A. Ziegler, "Recent developments in the analysis of carbon in iron and iron alloys," Trans. Am. Electrochem. Soc. 56, 231–238 (1929).

[9] L. A. Wooten and W. G. Guldner, "Determination of carbon in low carbon iron and steel," Ind. Eng. Chem. 14, No. 10, 835 (1942).

carburization stages can generally be observed. Cross sections of filaments decarburized in a vacuum by high temperature flashing showed decarburized surface layers.

Referring to the structure of the normal filament as reproduced in Fig. 1, it can now be seen that this type of carbide formation was probably produced by decarburization. The hydrocarbon content of the hydrogen gas used during carburization was found to be too high for direct formation of laminated material. It follows, therefore, that decarburization must have taken place during the exhaust cycle when the filament was heated to a high temperature. In practice the duration of this cycle was found to be determined by the speed with which a certain degree of vacuum was obtained. This means that the filament was heated in the presence of gas and that the time varied with circumstances determining the degassing speed of the tube parts. Varying degrees of decarburization can easily be explained this way. If, in addition to that, the carbide formation during carburization does not produce constant results, as was found caused by improper dosing of the hydrocarbon in the hydrogen, it follows that a wide variety of carbide structures must be expected.

Varying amounts of carbide, of course, would cause differences in the values of cold resistance and of filament current (hot resistance). But why should two filaments having the same values of cold resistance have different values of filament current? This phenomenon was theoretically explained as follows: The cold resistance of a filament depends upon the amount of $W_2C$ formed, but it does not matter where in the filament this $W_2C$ is present. It may be on the surface as in the case of a freshly carburized filament; or it may be under the surface as in the case of a decarburized filament. Therefore, two filaments can have the same cold resistance, but different surface conditions.

TABLE I.

| Weight of sample in grams | % carbon in sample | % carbon in shell | |
|---|---|---|---|
| 0.0200 | 0.350 | 2.45 | |
| 0.0161 | 0.403 | 2.31 | |
| 0.0136 | 0.538 | 2.45 | Average |
| 0.0141 | 0.534 | 2.36 | 2.45 |
| 0.0155 | 0.508 | 2.69 | |
| 0.0258 | 0.400 | 2.45 | |

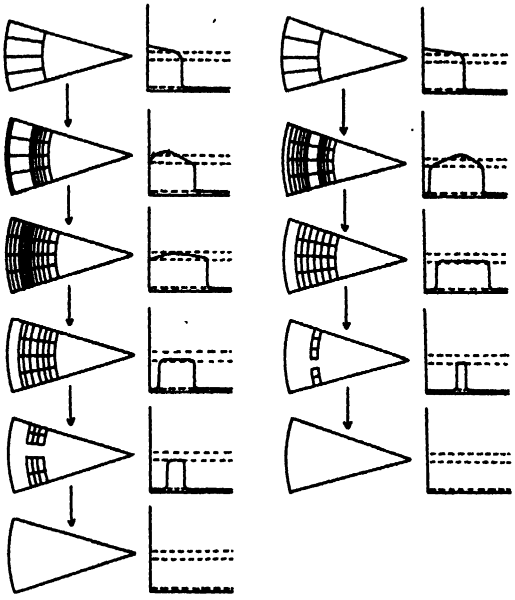


FIG. 11. Series showing effects of progressive decarburization: a, in dry hydrogen; b, in wet hydrogen.

When, however, these filaments are both put on the same operating voltage, as is done during tube operation, the one with the tungsten surface, having a lower heat emissivity, will have a higher temperature, a higher hot resistance, and a lower filament current than the one with the tungsten carbide surface, which has a higher heat emissivity.

The fact that surfaces intermediary between almost pure tungsten and total tungsten carbide as extremes may occur, due to decarburization on exhaust, opens the possibility of a range of hot resistances, corresponding to each cold resistance value. Thus, the cold resistance-*versus*-filament current diagram becomes an area instead of a line.

For confirmation of this theory, special tubes were built containing two identical tungsten filaments having the same initial cold resistance. One of the filaments was carburized under circumstances which guaranteed the formation of the massive $W_2C$ phase. The other filament was overcarburized and, subsequently, decarburized in wet hydrogen to exactly the same cold re-

TABLE II.

| Carbide-surface filament | | | | Tungsten-surface filament | | | |
|---|---|---|---|---|---|---|---|
| $V$ | $A$ | $VA$ | $T$°C | $V$ | $A$ | $VA$ | $T$°C |
| 7.5 | 2.08 | 15.6 | 1528 | 7.5 | 1.9 | 14.25 | 1660 |
| 10.0 | 2.37 | 23.7 | 1725 | 10.0 | 2.21 | 22.1 | 1815 |
| 12.5 | 2.63 | 32.9 | 1904 | 12.5 | 2.54 | 31.75 | 1960 |
| 14.3 | 2.8 | 40.0 | 1970 | 14.3 | 2.70 | 38.6 | 2000 |
| 15.3 | 2.9 | 44.4 | 2000 | | | | |

sistance as the first filament. When these two filaments were operated in series, the decarburized filament showed a visibly higher temperature.

Measurements made on the separate filaments are given in Table II.

These data show that the purposely decarburized filament has the lowest filament current and the highest temperature at a fixed operating voltage. In this case the difference in filament current values at 2000°C amounts to about 4 percent.

If we were to operate the filaments at 14.3 volts, the temperature difference is 30°C. Furthermore, these data agree with Barnes' results,[5] *i.e.* that the total emissivity of tungsten carbide is higher than that of tungsten for the same temperature of operation.

This is an appropriate place to return to our first question as to why the thyratron control sometimes failed to produce filaments of constant cold resistance, because the answer was also found to be differences in surface conditions. A filament with a surface having total emissivity higher than average will produce a lower than normal temperature for normal input. Translated into terms of hot resistance this means that a lower temperature corresponds to the hot resistance at which the thyratron relay is set to act. This in turn requires that overcarburization must take place if the necessary hot resistance is to be reached. A filament of higher than normal cold resistance is the result.

Two cases of abnormally high emissivity of filaments were encountered. The first one occurred when helical filaments were oxidized heavily during the forming operation and then were reduced by flashing them in a hydrogen atmosphere. This procedure produced a very rough surface. In the second case, the hydrocarbon content of the carburizing atmosphere was so great that a carbon deposit was formed on the surface of the filament.

From the above investigations, the following conclusions can be drawn:

1. Thyratron control for the carburization of thoriated-tungsten filaments is an excellent method, provided the filaments to be carburized have uniform surface conditions, and provided a strict control of the hydrocarbon content in the hydrogen atmosphere is maintained.
2. The character of the carbide formations obtained during carburization and decarburization can be explained on the basis of the theory of diffusion and the characteristics of the carbon-tungsten phase diagram.
3. The amount of decarburization which occurs during tube manufacture has an important effect on the uniformity of the electrical filament characteristics of tubes.
4. The hot-to-cold resistance ratio of a filament is greatly influenced by the surface condition of the filament.

# Creep Deflections in Columns

JOSEPH MARIN*
*School of Engineering, The Pennsylvania State College, State College, Pennsylvania*

(Received August 22, 1946)

The determination of column deflections and column buckling loads has been considered for many years. The available theories, however, do not provide for materials which creep with time under constant loads. For the design of structural members made of these materials, a consideration of creep may be of practical importance. Plastics, concrete, and some metals creep at normal temperatures while other metals creep only at high temperatures and at stress values beyond the yield point. A consideration of creep in the design of some structures appears appropriate in view of the modern developments in plastics and the presence of high stress values which are sometimes beyond the yield stress. This paper gives a rational theory for predicting creep deflections in columns. A special case using this theory is applied to the interpretation of some preliminary tests of an aluminum alloy.

## INTRODUCTION

THE usual concept of failure in columns by elastic instability must be modified if the material creeps. For columns made of materials that creep, critical buckling loads are smaller than the elastic buckling loads since the progressive lateral creep deflections must be considered. It is necessary therefore to modify our definition of the critical buckling load and to define the buckling load as that load which will produce a large lateral deflection at the end of a specified time. To determine these loads, a theory will be developed giving the creep deflections in a column in terms of the load, column dimensions, material constants, and time. These theoretical deflections will then be compared with the corresponding experimental values.

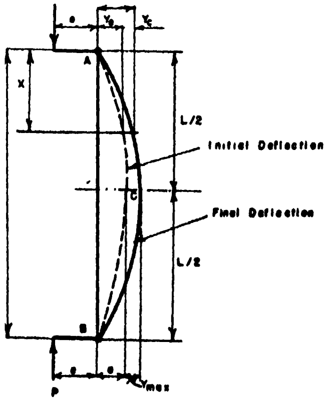


FIG. 1. Pin-ended column subjected to an eccentric load.

## GENERAL THEORY FOR COLUMN DEFLECTIONS

Columns having cross sections with at least one axis of symmetry will be considered. The column $AB$ is free at the upper end $A$, and at the base $B$, and is loaded eccentrically with a load $P$ in the plane of symmetry (Fig. 1). Initially, an elastic lateral deflection $y_e$ is produced at a distance $x$ from point $A$. This initial deflection will include the elastic deflection plus the initial creep deflection as shown in Fig. 2. Actually the column deflects so that $AB$ rather than $AC$ represents the initial elastic and creep deformation and beyond $B$ for a short period of time the deflections are represented by the curve $BD$. Beyond point $D$ the experimental deflection-time relation $DE$ was found to be linear. It will be assumed in this theory that the actual deflection curve $ABDE$ is replaced by $ACE$. To determine the lateral deflection ($y=y_e+y_c$) at a time $t$, a theory for defining

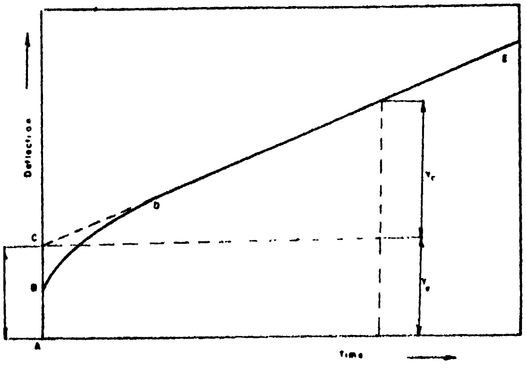


FIG. 2. Deflection-time relation.

---

* Professor of Engineering Mechanics.

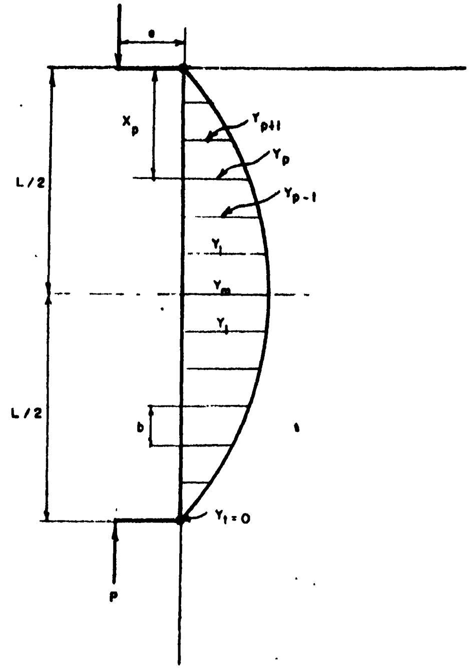


FIG. 3.

creep deflection in bending corresponding to the elastic theory for deflection must first be obtained.

A theory for the creep deflection in bending has been proposed by the writer in the form of a differential equation for the creep deflection $y_c$ in terms of the moment $M$ and time $t$.[1] It was shown by this theory that

$$(D/t)(d^2y_c/dx^2)=M^n. \qquad (1)$$

In Eq. (1), $D$ corresponds to the flexural rigidity $EI$ in the elastic theory. Its value for a rectangular cross section is

$$D=\frac{1}{A}\frac{(2b)^n(h/2)^{2n+1}}{(2+1/n)^n}, \qquad (2)$$

where $A$ and $n$ = experimental creep constants,

$b$ = the width of the section,
$h$ = the depth of the section.

Equation (1) is based on the log-log creep stress law $C=AS^n$ where $C$ is the assumed constant rate of creep strain and $S$ is the unit stress.

The moment $M$ in terms of the load $P$ at a time $t$ is, from Fig. 1,

$$M=-P(e+y)=-P(e+y_e+y_c), \qquad (a)$$

where $e$ = the eccentricity of the load. Substituting the value of the moment $M$ from Eq. (a) in Eq. (1), the differential equation for the creep deflection $y$ at a distance $x$ is

$$(D/t)(d^2y_c/dx^2)=-P^n(e+y_e+y_c)^n. \qquad (3)$$

In this equation $x$, $y_e$, and $y_c$ are variables. These three variables can be reduced to the necessary two by expressing the initial deflection $y_e$ in terms of $x$. For this purpose the shape of the curve $AB$ can be assumed with reasonable accuracy. This assumption is similar to that used in the analysis of buckling problems in the elastic case using the Rayleigh method. For the present problem the shape of the curve $AB$ will be assumed as a sine function. That is, the value of $y_e$ in terms of $x$ will be assumed as

$$y_e=a\sin\frac{\pi x}{L}, \qquad (b)$$

where $a$ = the initial deflection at point $C$ in the column and $L$ = the length of the column.

Placing the value of $y_e$ from Eq. (b) in Eq. (3) and noting that $y=y_e+y_c$, the differential equation for the total deflection $y$ as a function of the distance $x$ becomes

$$\frac{d^2y}{dx^2}=-\frac{a\pi^2}{L^2}\sin\frac{\pi x}{L}-\frac{P^nt}{D}(e+y)^n. \qquad (4)$$

The values of $n$ in Eq. (4) must be assumed as odd integers since the creep stress law $C=AS^n$ must apply for compressive and tensile values of $S$ in the case of bending where both types of stress exist.[2] That is, Eq. (4) must be solved for values of $n=1, 3, 5\cdots n$, where $n$ is an odd number. A direct solution of Eq. (4) for $y$ in terms of $x$ could not be found. One method of attack using numerical integration is given in Appendix 1. A general method of solution is developed in this paper by using a trial and error

---

[1] J. Marin, "Creep of aluminum subjected to bending at normal temperature," Proc. A.S.T.M. 40, 937 (1940); and "Methods of structural analysis in the case of creep," J. West. Soc. Eng. 45, 301 (Dec. 1940).

[2] J. Marin, "Mechanics of creep for structural analysis," Trans. A.S.C.E. Vol. 108, 470 (1943).

procedure. This method is similar to the one used by Holzer[3] for the determination of natural frequency of vibration of multi-mass systems. The procedure used is to replace the differential Eq. (4) by an equivalent system of difference equations. To do this, consider the column divided into a number of segments $N$ of length $b=L/N$ (Fig. 3). The ordinates to the deflection curve will be designated by $y_m, y_1 \cdots y_p$, as shown in Fig. 3. The second derivative of $y$ with respect to $x$, $d^2y/dx^2$, at any section where $x=x_p$, is given approximately by the equation:

$$\left(\frac{d^2y}{dx^2}\right)_{x=x_p}=\frac{1}{b^2}(y_{p+1}-2y_p+y_{p-1}). \tag{5}$$

Placing the value of $d^2y/dx^2$ from Eq. (4) in Eq. (5) and solving for $y_{p+1}$,

$$y_{p+1}=2y_p-y_{p-1}-F_p+Q(e+y_p)^n, \tag{6}$$

where

$$F_p=(b^2a\pi^2/L^2)\sin(\pi x_p/L), \tag{7}$$

and

$$Q=b^2P^nt/D. \tag{8}$$

Denoting the deflection at the mid-section of the column by $y_m$ and at the end by $y_t$ and writing Eq. (6) for each section successively,

$$\left.\begin{aligned} y_1&=y_m-\tfrac{1}{2}[F_m+Q(e+y_m)^n],\\ y_2&=2y_1-y_m-[F_1+Q(e+y_1)^n],\\ y_3&=2y_2-y_1-[F_2+Q(e+y_2)^n],\\ y_{p+1}&=2y_p-y_{p-1}-[F_p+Q(e+y_p)^n],\\ y_t&=2y_{t-1}-y_{t-2}-[F_{t-1}+Q(e+y_{t-1})^n],\end{aligned}\right\} \tag{9}$$

where

$$F_m=\frac{b^2a\pi^2}{L^2}\sin\frac{\pi x_m}{L},\quad F_1=\frac{b^2a\pi^2}{L^2}\sin\frac{\pi x_1}{L}. \tag{10}$$

The method of applying Eqs. (9) to determine column deflections for a given load $P$ at a time $t$ is as follows:

1. From the column dimensions, material constants $A$ and $n$, time $t$, and load $P$, determine the value of $Q$ by Eq. (8).
2. Using the value of the elastic deflection for the approximate value of $a$ and for different values of $x$ determine the values of $F_1, F_2 \cdots F_p$.
3. Assume a value for $y_m$ and, by Eqs. (9), determine successively values of $y_1$ to $y_t$. The value of $y_t$ should be zero if the correct value of $y_m$ was assumed.
4. Repeat Item 3 until small positive and small negative values of $y_t$ are obtained. Then by interpolation the correct value of $y_m$ can be calculated.

The above calculations can be repeated for various values of $P$ and $t$. These data can then be used to determine the load deflection curves with the time $t$ as a parameter (Fig. 4).

For the aluminum alloy tested in this investigation it was found that $n=1$ is a good approximation. As shown in the following, the solution of Eq. (3) is greatly simplified when $n=1$.

## SPECIAL THEORY FOR COLUMN DEFLECTIONS WHEN $n=1$

By Eqs. (3) and (b) for $n=1$,

$$\frac{D}{t}\frac{d^2y_c}{dx^2}=-P\left(e+a\sin\frac{\pi x}{L}+y_c\right), \tag{c}$$

or

$$\frac{d^2y_c}{dx^2}+k^2y_c=-k^2a\sin\frac{\pi x}{L}-k^2e, \tag{11}$$

where $k^2=Pt/D$.

It can be shown that the solution of this differential equation is

$$y_c=C_1\sin kx+C_2\cos kx+\frac{a\sin\pi x/L}{\dfrac{\pi^2}{k^2L^2}-1}-e. \tag{12}$$

Boundary conditions can be used to determine the constants $C_1$ and $C_2$. Since $y_c=0$ when $x=0$, from Eq. (12),

$$C_2=e. \tag{d}$$

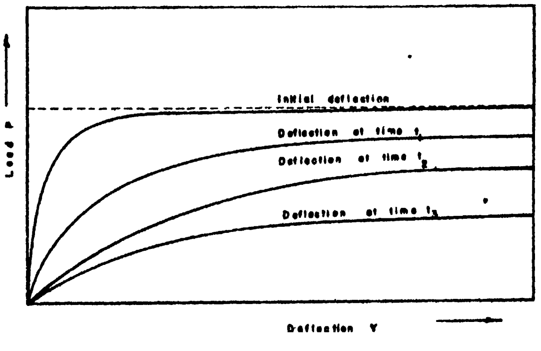


FIG. 4. Load-deflection relations.

---

[3] J. P. Den Hartog, *Mechanical Vibrations* (McGraw-Hill Book Company, Inc., New York, 1940), p. 225. See also Dana Young, "Inelastic buckling of variable section columns," paper presented at the Annual Meeting of the A.S.M.E., Dec. 1944.

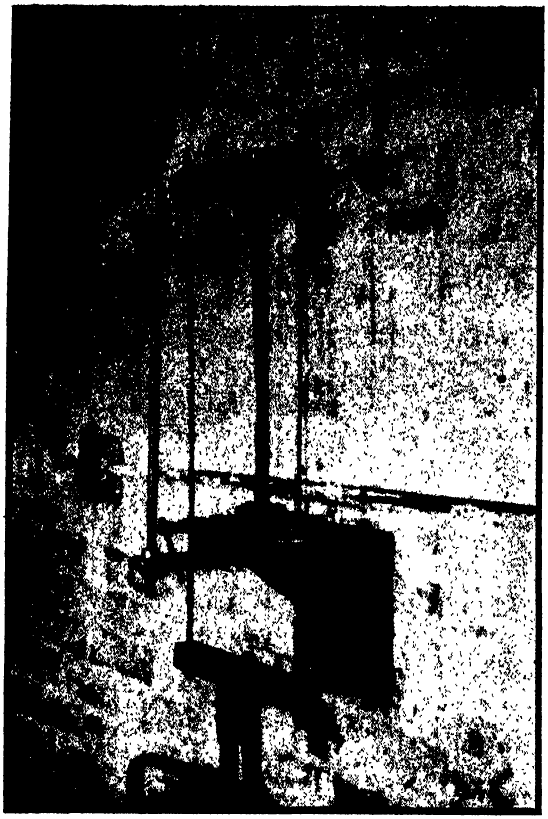


FIG. 5. Creep buckling apparatus.

The deflection $y_e$ also equals zero when $x=L$, or from Eq. (12),

$$C_1=\frac{e(1-\cos kL)}{\sin kL}. \tag{e}$$

Substituting the values of $C_1$ and $C_2$ from Eqs. (d) and (e) in Eq. (12) the creep deflection is

$$y_e=\frac{\sin kx}{\sin kL}(1-\cos kL)e+e\cos kx-e+\frac{a\sin\pi x/L}{\dfrac{\pi^2}{k^2L^2}-1}. \tag{13}$$

By inspection the maximum creep deflection is for a value of $x=L/2$, or

$$y_{\max}=e\left(\sec\frac{PtL^2}{4D}-1\right)+\frac{aPtL^2/D}{\pi^2-\dfrac{PtL^2}{D}}. \tag{14}$$

If it is assumed that the initial deflection $y_e$ is entirely the elastic deflection at the center of the column, then it can be shown by the usual elastic theory that,

$$y_e=e\sec[(PL^2/4EI)]^{\frac{1}{2}}-e. \tag{15}$$

The total deflection at the center ($y_T$) is equal to the creep deflection ($y_{\max}$), as given by Eq. (14), plus the elastic deflection, as given by Eq. (15). That is, by Eqs. (14) and (15),

$$y_T=e\left[\sec\frac{PtL^2}{4D}+\sec[(PL^2/4EI)]^{\frac{1}{2}}-2\right]+\frac{aPtL^2/D}{\pi^2-\dfrac{PtL^2}{D}}. \tag{16}$$

The maximum moment produced in the column at a time $t$ is,

$$M_{\max}=P(y_T+e). \tag{17}$$

The maximum bending stress with $n=1$ is the same as for the elastic case or equal to $M_{\max}c/I$, where $I=$ the moment of inertia of the cross section and $c=$ the distance to the outer fiber. In addition to the bending stress, there is an axial compressive stress equal to the load divided by the area or $P/A$. Then the maximum stress is

$$S_{\max}=\frac{P}{A}+\frac{M_{\max}c}{I}. \tag{18}$$

Using values of $M_{\max}$ and $y_T$ from Eqs. (16) and (17) in Eq. (18), the maximum stress is

$$S_{\max}=\frac{P}{A}+\left(\frac{Pc}{I}\right)\left[e\sec(PtL^2/4D)+e\sec[(PL^2/4EI)]^{\frac{1}{2}}-e+\frac{aPtL^2/D}{(\pi^2-PtL^2/D)}\right]. \tag{19}$$

TABLE I. Physical properties of 3S-H aluminum alloy.

| | |
|---|---|
| Yield strength (0.2 percent set) (pounds per square inch) | 25,000 |
| Ultimate strength (pounds per square inch) | 29,000 |
| Elongation (percent in 2 inches) (one-half inch round specimen) | 10 |
| Hardness (Brinell—500 Kg) | 55 |
| Modulus of elasticity (pounds per square inch) | 10,300,000 |
| Modulus of rigidity (pounds per square inch) | 3,800,000 |

Equation (19) corresponds to the secant formula obtained in the elastic theory and it can be used to determine maximum stresses in columns accompanied by creep.

## EXPERIMENTS ON ECCENTRICALLY LOADED COLUMNS

The columns tested were supported and loaded as shown in Fig. 5. The load was applied to the column specimens by a hanger supplied with a universal joint to prevent bending, and provision was made for rigidly fixing the lower end of the column. The lateral creep deflections were measured by a Federal dial reading to 1/10,000 inch. Precautions were taken for placing the column initially vertical and the dial pointer horizontal. The material tested was an aluminum alloy designated as 3S–H by the Aluminum Company of America and the properties of this material, as supplied by Mr. R. L. Templin, Chief of Tests of this Company, are given in Table I.

Tension and pure bending creep tests also were made on this material at room temperatures.[1] The tension creep tests showed that the creep rate-stress relation was a straight line on a log-log plot. That is, the variation between the stress $S$ and constant creep rate $C$ is:

$$C = AS^n, \tag{20}$$

where $A$ and $n$ are experimental constants.

The values of $A$ and $n$ as determined by the tension tests were reported[1] as $A = 6.51 \times 10^{-12}$ and $n = 1.18$. A reconsideration of this data shows that a value of $n = 1$ and $A = 2.5 \times 10^{-11}$ is also a good approximation. The value of one for the constant $n$ is selected since this simplifies the theoretical interpretation of the test results. The pure bending creep tests showed that Eq. (1) gives a reasonably good approximation for this material.

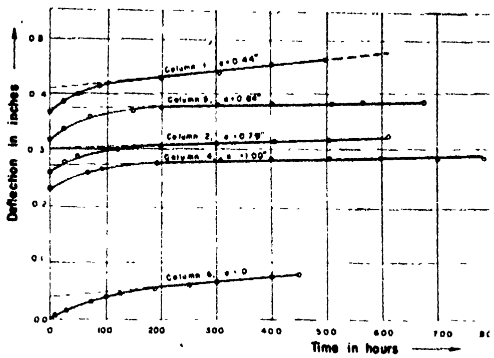


FIG. 6. Deflection-time relations.

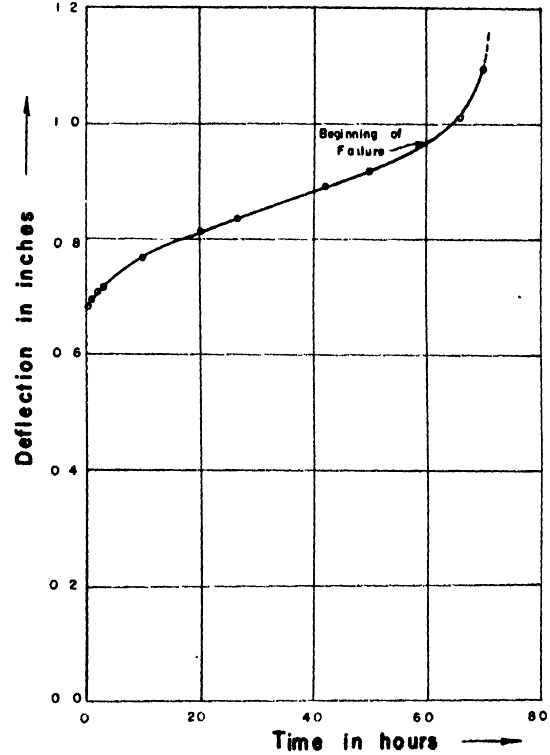


FIG. 7. Creep deflection-time curve for column 3.

TABLE II. Dimensions and loads for columns.

| Col. No. | Length ($L/2$) in. | Width ($b$) in. | Thickness ($h$) in. | Slenderness ratio ($L/r$) | Eccentricity ($e$) in. | Load ($P$) (lb.) | Moment ($M$) in. lb. |
|---|---|---|---|---|---|---|---|
| 1 | 14.0 | 0.50 | 0.49 | 195 | 0.44 | 258 | 113.5 |
| 2 | 14.0 | 0.50 | 0.50 | 195 | 0.79 | 143 | 113.0 |
| 3 | 14.0 | 0.50 | 0.49 | 195 | 0.25 | 308 | 77.0 |
| 4 | 14.0* | 0.50 | 0.48 | 195 | 1.00 | 113 | 113.0 |
| 5 | 14.0 | 0.50 | 0.49 | 195 | 0.64 | 177 | 113.3 |
| 6 | 14.0 | 0.50 | 0.40 | 195 | 0. | 433 | 0. |
| 7 | 10.2 | 0.50 | 0.36 | 213 | 0.34 | 158 | 53.8 |
| 8 | 10.2 | 0.50 | 0.36 | 213 | 0.38 | 158 | 60.0 |

The column specimens used were about one-half inch square, and most of them were about 15 inches in length. To prevent local yielding the bases of the columns were increased in size. Various values of the eccentricity $e$ and load $P$ were used for the eight columns tested, as is shown in Table II. Table II also lists the average dimensions and slenderness ratios for the mem-

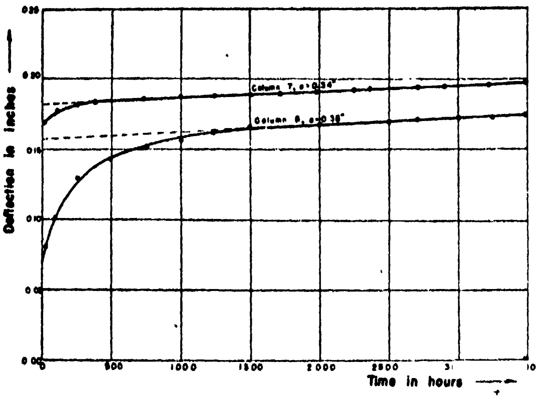


FIG. 8. Deflection-time relations for columns 7 and 8.

bers used. For each experiment, lateral deflections were measured at a point near the upper end of the column, as shown in Fig. 5. An approximate value of the initial elastic and creep deflections was noted for each member. Creep deflection readings were recorded at varying intervals of time and deflection-time plots are shown for the various specimens in Figs. 6–8. Each test covered a time period of at least 500 hours and two specimens covered a time of 3000 hours. After the first 24 hours the creep rate became almost constant so that readings were taken only once a day. The creep rate was constant for all columns except No. 3 which failed by buckling at the end of 70 hours (Fig. 7). For a number of the specimens tested the applied moment was maintained constant by decreasing the load and increasing the eccentricity.

## INTERPRETATION OF TEST RESULTS

A comparison between the theoretical and experimental deflection values can be made by determining both the creep deflections and creep rates using each method. The creep deflection given by Eq. (14) also represents the maximum theoretical creep deflection for the column of Fig. 5. For example, to determine the theoretical creep deflection at 500 hours in the case of column 2, it is necessary to substitute in Eq. (14) the values of $e=0.79$ inches, $P=143$ lb., $t=500$ hr., $L=28$ in., $D=I/A=2.08\times 10^8$, and $a=0.364$ in. Then the theoretical creep deflection $y_{max}=0.042$ in. The values of the theoretical deflections for other columns are given in Table III.

The actual deflections were not measured at the ends of the columns so that it is necessary to correct these deflections in order to compare them with the theoretical values. To do this the creep deflection curve is assumed to be a sine curve. The corrected values of the measured creep deflections are listed in Table III for comparison with the theoretical values. The values of the creep rates, as obtained by theory and experiment, are also given in Table III for purposes of comparison.

## CONCLUSION

In view of the preliminary nature of these tests and the approximations made in the theory, the agreement between the theoretical and experimental creep values as given in Table III is quite satisfactory. There is need, however, for further research on this problem before a definite conclusion can be reached.

## ACKNOWLEDGMENTS

The author wishes to express his thanks to R. L. Templin, Chief of Tests of the Aluminum Company of America for the aluminum used in these tests and for control test data supplied. The experimental data of this paper was taken from a Master's Thesis by R. C. DeHart performed under the direction of the writer.

## APPENDIX 1

An alternate method for the solution of Eq. (1) is given in the following. The value of $y_e$ is first assumed to be $y_e=ix^2$. The constant $i$ can be determined by noting that $y_e=a$ when $x=L$ (Fig. 9) or $i=a/L^2$. That is,

$$y_e=(a/L^2)x^2, \tag{f}$$

TABLE III. Comparison of theoretical and experimental values.

| Column number | Moment (in. lb.) | Eccentricity (inches) | Creep deflec. (in.) Exp. | Creep deflec. (in.) Theor. | Creep rates (in./in./hr.$\times 10^4$) Exp. | Creep rates (in./in./hr.$\times 10^4$) Theor. |
|---|---|---|---|---|---|---|
| 1 | 113.5 | 0.44 | 0.060 | 0.057 | 1.20 | 1.14 |
| 2 | 113.0 | 0.79 | 0.040 | 0.042 | 0.80 | 0.84 |
| 4 | 113.0 | 1.00 | 0.030 | 0.038 | 0.060 | 0.076 |
| 5 | 113.3 | 0.34 | 0.031 | 0.043 | 0.062 | 0.086 |
| 7 | 53.8 | 0.34 | 0.018 | 0.029 | 0.060 | 0.097 |
| 8 | 60.0 | 0.38 | 0.018 | 0.037 | 0.060 | 0.123 |

*Note:* For columns 1 to 5 creep deflections are for $t=500$ hours. For columns 7 and 8 creep deflections are for $t=3000$ hours.

where $a$ = the initial deflection at $B$ and $L$ = the length of the column. For the column in Fig. 9 the moment is

$$M=Py'=P(e+y_{\max}+a-y_c-y_e). \qquad \text{(g)}$$

Placing the value of $y_e$ from Eq. (f) and the value of $M$ from Eq. (g) in Eq. (1),

$$d^2y_c/dx^2=(k_1y_c+k_2x^2+k_3)^n, \qquad (21)$$

where

$$k_1=-P(t/D)^{1/n}, \quad k_2=\frac{-aP}{L^2}(t/D)^{1/n},$$

and

$$k_3=(e+y_{\max}+a)P(t/D)^{1/n}.$$

Placing

$$u=k_1y_c+k_2x^2+k_3 \qquad \text{(h)}$$

in Eq. (18),

$$d^2u/dx^2=k_1u^n+2k_2. \qquad \text{(i)}$$

Integrating Eq. (i) and separating the variables,

$$\int\frac{du}{(u^{n+1}+k_4u+k_5)^{1/2}}=k_6\int dx, \qquad (22)$$

where

$$k_4=(2k_2/k_1)(n+1),$$
$$k_5=-(k_3)^{n+1}-(n+1)(2k_2k_3/k_1),$$
$$k_6=[(n+1)/(2k_1)]^{1/2}.$$

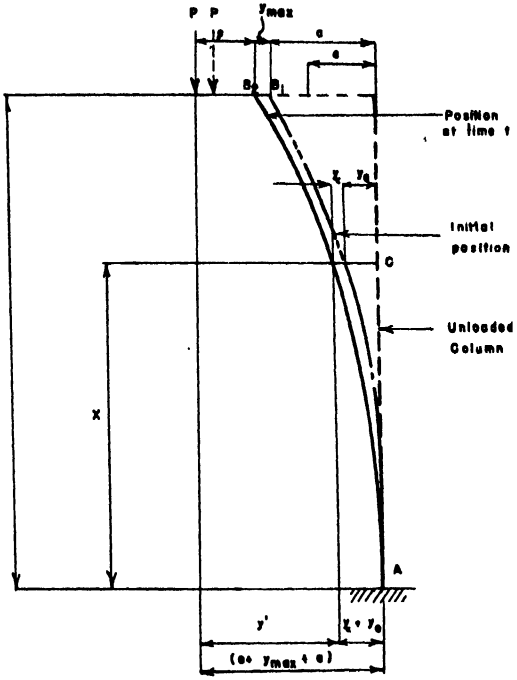


FIG. 9. Deflection of column fixed at one end and free at the other.

Equation (22) can now be solved for particular values of $n$ by using numerical integration methods. For a value of $n=1$, Eq. (22) can be integrated directly, or

$$u^2+k_4u+k_5=(e^{k_6x+k_7}-x-k_4/2)^2, \qquad (23)$$

where $k_7$ is a constant of integration.

Replacing $u$ by its value in terms of $x$ and $y$ as given by Eq. (h), the equation for the deflection $y_c$ can be obtained.

# Effects of Magnetic Field on Oscillations and Noise in Hot-Cathode Arcs*

J. D. Cobine and C. J. Gallagher**
*Radio Research Laboratory, Harvard University, Cambridge, Massachusetts*

(Received April 11, 1946)

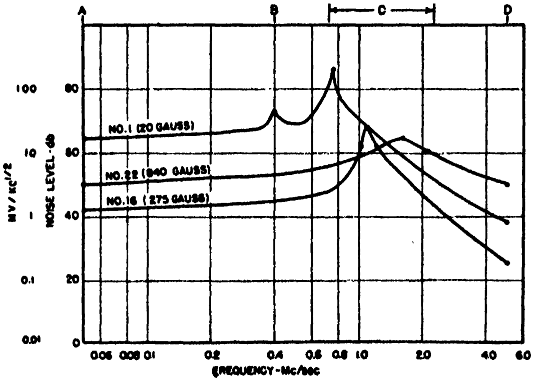
The voltage disturbances in hot-cathode arcs are markedly affected by a magnetic field transverse to the normal flow of current. The normal plasma oscillations can be suppressed and the noise level above 1 Mc/sec. greatly increased. The oscillations appear to be transmitted to the electrodes by electrons whose velocity is modulated by the plasma ion oscillations as they travel directly from cathode to anode. The effect of the magnetic field is to remove these electrons from the current reaching the anode. A critical value of magnetic field will suppress the oscillations and reduce the noise to a minimum. As the magnetic field is further increased the high frequency noise level increases. The high frequency noise level also increases if the pressure decreases, or if a lighter gas is used. The noise level is independent of the plasma volume. The effect of the magnetic field is to build up a region of high ion density in a narrow region near the electron source. Plasma oscillations with continuous spectrum are generated in this region of high density and high density gradient.

## INTRODUCTION

Occasionally mention is made of the fact that gas discharge tubes are "noisy," where "noise" is taken to mean any objectionable fluctuating voltage, such as random noise or oscillations.

Many investigators have reported the existence of oscillations in the voltage appearing between the electrodes of gaseous discharge.[1] The oscillations so observed appear to be of three types, *viz.*, oscillations of electrons and of positive ions in the plasma, and relaxation oscillations. The former appear to be characteristic of the discharge (current, pressure, geometry, nature of gas), while the latter are dependent on both the discharge and the external circuit conditions. Observations and analysis of the plasma oscillations have been made by Tonks and Langmuir.[2] These workers showed theoretically that the frequency of the plasma oscillations increases with increasing positive-ion density. The frequencies of these oscillations have been studied in narrow ranges from 1000 kc/sec to 1000 Mc/sec.

No systematic study of the random noise developed by gaseous discharges has been reported in the literature although gas tubes have been used as random noise generators in acoustical testing,[3] and Tonks and Langmuir[2] mention the existence of a continuum in the voltage output of their discharge tube. Recently the authors have reported a study of random noise and oscillations in hot-cathode arcs.[4] In this study it was shown that the random noise could be explained as the result of random fluctuations of space charge at the electrodes.



Anode current = 30 ma
Load resistance = 2500 ohms

Fig. 1. Selected spectra of hot-cathode gas triode (RCA 884) with transverse magnetic field.

---

* This work was done in whole or in part under Contract No. OEMsr-411 between the President and Fellows of Harvard College and the Office of Scientific Research and Development.

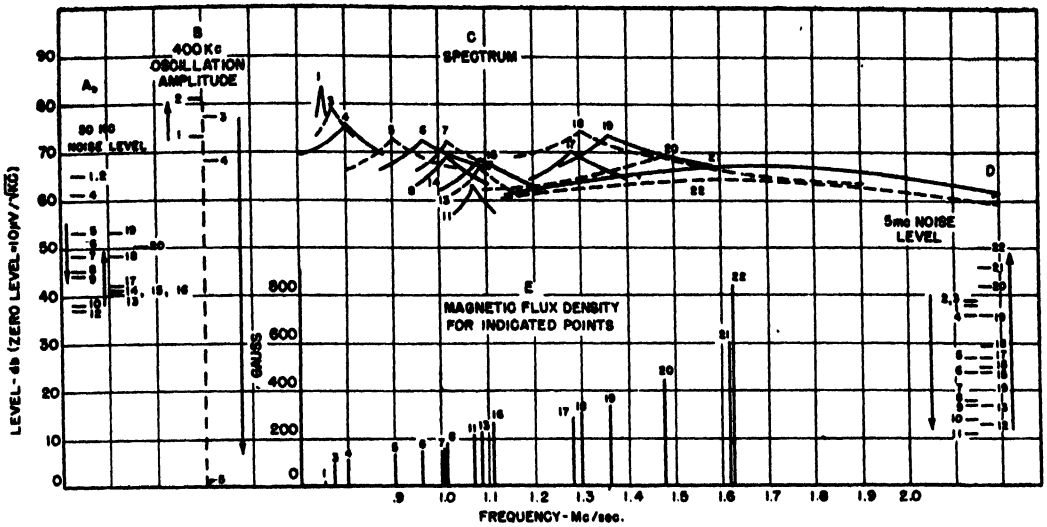
** Both authors now at Research Laboratory, General Electric Company, Schenectady, New York.

[1] E. V. Appleton and A. G. D. West, Phil. Mag. **45**, 879 (1923). C. Eckart and K. T. Compton, Phys. Rev. **24**, 97 (1924). F. H. Newman, Phil. Mag. **47**, 839 (1924). K. T. Compton and C. Eckart, Phys. Rev. **25**, 139 (1925). R. E. Clay, Phil. Mag. **50**, 985 (1925). L. A. Pardue and J. S. Webb, Phys. Rev. **32**, 946 (1928). E. W. B. Gill, Phil. Mag. **8**, 955 (1929). H. Kniepkamp, Zeits. f. Tech. Phys. **17**, 397 (1936). W. Funk and R. Seeliger, Zeits. f. Physik **113**, 203 (1939).

[2] I. Langmuir, Proc. Nat. Acad. Sci. **14**, 627 (1928); L. Tonks and I. Langmuir, Phys. Rev. **33**, 195 (1929).

[3] E. Peterson, Bell Lab. Record **18**, 81 (1939). H. Dudley, R. R. Riesz, and S. S. A. Watkins, J. Frank. Inst. **227**, 739 (1939).

[4] J. D. Cobine and C. J. Gallagher, Phys. Rev. **70**, 113 (1946). J. Frank. Inst. **243** (Jan. 1947).



Anode current = 30 ma
Load resistance = 2500 ohms
A—50 kc/sec. noise levels
B—Amplitude of 400 kc/sec. oscillation
C—Spectra, peak and points 6 db down
D—5 mc/sec. noise levels
E—Magnetic flux density for parts A, B, C, D
Reference numbers correspond to magnetic flux densities given in section E

FIG. 2. Effect of magnetic field on noise and oscillations of hot-cathode gas triode (RCA 884).

The present paper presents a further study of noise and oscillations in gas discharges when placed in a magnetic field. A transverse magnetic field is shown to produce two new effects, namely, the suppression of oscillations and the complete alteration of the noise spectrum.

## EXPERIMENTAL TECHNIQUE

The voltage fluctuations can be studied effectively by means of spectrum analyzers which measure amplitude and frequency characteristics of the oscillations and random noise. The frequency range from 25 c.p.s. to 9 Mc/sec. could be continuously studied with three heterodyne-type analyzers having band widths sufficiently small that the continuous noise spectrum remained essentially uniform over the region under observation.[5] Thus, the analyzers measured the r.m.s. voltage over a small band of frequencies. Since the r.m.s. noise power in a uniform spectrum is proportional to the band width, all measurements could be referred to a common band width, arbitrarily chosen as 1 kc/sec. The measurements of the spectra under various conditions are presented in db referred to a zero db level of $10\mu v/(\text{kc/sec.})^{\frac{1}{2}}$. Precautions were taken to insure that the input impedance of each analyzer was sufficiently high to avoid "loading" the gas tube output circuit and thus distorting the spectrum being measured. The discharge tube being studied was connected in series with a non-inductive resistance and a regulated power supply. The fluctuation voltage was measured directly between the tube electrodes. The data presented were obtained from both commercial tubes and tubes of experimental design.

## EXPERIMENTAL RESULTS

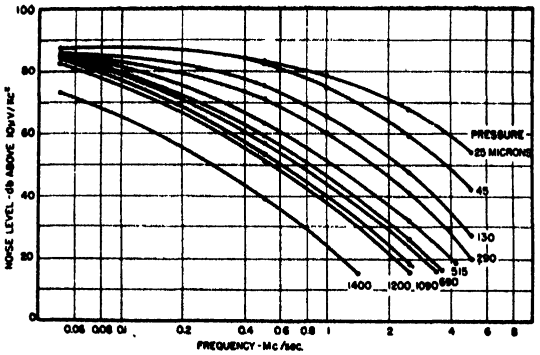
### Oscillations

Much information about the oscillations was conveniently obtained by studying characteristics of type 884 tubes (hot-cathode, argon-filled gas triode). As has been shown previously,[4] the voltage between anode and cathode contains two oscillations superimposed on a background of random noise. These oscillations are not harmonically related.

When a magnetic field is applied transverse to the normal direction of current flow the effect on each oscillation is quite different as shown in Figs. 1 and 2. The effect of the magnetic

---

[5] G. P. McCouch and P. S. Jastram, "Video Spectrum Analyzers," Radio Research Lab. Rep., No. 411-96. P. S. Jastram, "Low-Frequency Spectrum Analyzer," Radio Research Lab. Rep., No. 411-154. J. D. Cobine and J. R. Curry, Rev. Sci. Inst. 17, 190 (1946).



Anode current = 10 ma
Magnetic flux density = 860 gauss

FIG. 3. Effect of pressure on spectrum of argon tube.

field on the lower frequency oscillations is indicated in Section B of these figures, where it is seen that as the magnetic field increases, the amplitude first increases and then decreases without any change in frequency. With sufficiently high field (approx. 150 gauss), this oscillation disappears completely. Oscillation C undergoes a more complicated series of changes. The peak shifts toward higher frequencies, with marked broadening. Figure 2, Section C, shows the amplitude of the peak and its spectrum measured between frequencies whose amplitudes are 6 db below those of the peak.

For the strongest magnetic fields, the peak becomes so broad that a discrete oscillation cannot be detected. The numbers on the figure indicate the spectrum levels measured at the corresponding magnetic field in Section E. The behavior of these oscillations is typical of all the hot-cathode arcs studied.

The disappearance of these oscillations was studied in an experimental tube with an oxide-coated cathode and a small plane anode three inches from it. The cathode was partly surrounded by a metal shield to define an electron beam, the limits of which could be plainly seen at a few hundred microns pressure. Under these conditions the discharge space was filled with plasma and the two characteristic oscillations were observed in the voltage spectrum. When a magnetic field was applied so as to deflect the beam of electrons away from the anode, the low frequency oscillation disappeared. The oscillation reappeared at reduced amplitude if the beam was directed against the glass wall of the tube so that electrons coming directly from the cathode were reflected to reach the anode. It is not possible to eliminate this oscillation in any tube in which the electrons could reach the anode directly in spite of the magnetic field. Thus, this oscillation is a positive-ion plasma oscillation whose energy is transmitted to the external circuit by velocity modulation of the electron beam passing directly from cathode to anode. Tonks and Langmuir[2] have shown that the plasma oscillations have zero group velocity and cannot transmit energy directly to the external circuit.

Unlike the plasma oscillation, the second type of oscillation has been observed at all currents, even down into the "Townsend" discharge region where no plasma exists.[4] Similar oscillations have been observed by Ballantine[6] in vacuum tubes containing mercury as a residual gas (pressures less than 1 micron). Since these oscillations appear under conditions in which no plasma is present, as well as under plasma conditions, the most promising explanation appears to be that proposed by Ballantine, *i.e.*, the oscillations are those of positive ions in the potential minimum at the cathode. If these are "cathode" oscillations, the magnetic field has the effect on frequency that might be expected, since the electron density increases near the cathode due to the bending of the electron paths in the magnetic field. The increase in density would cause a deepening of the potential well, with consequent increase in oscillation frequency. The broadening of the peak is more difficult to explain.

## Noise

As shown in Fig. 1, the oscillations are superimposed on a continuous background of random noise. The effect of the magnetic field is first to

TABLE I

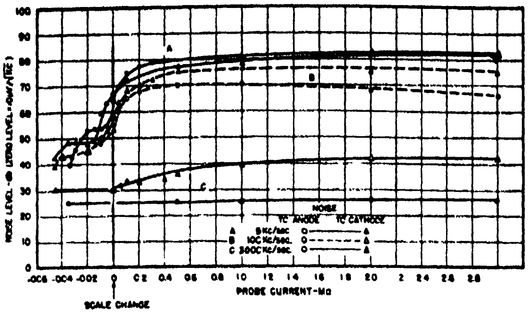
Noise characteristics of hot-cathode arcs in different gases
(1000 microns, 10 ma, 720 gauss)

| Gas | | He | A | Xe | Hg* |
|---|---|---|---|---|---|
| Noise level (db above 10$\mu$v/kc$^{\frac{1}{2}}$) | 50 kc/sec. | 89 | 73 | 83 | 80 |
| | 1 Mc/sec. | 81 | 69 | 63 | 62 |
| | 5 Mc/sec. | 48 | 42 | 32 | 33 |
| Peak-to-peak voltage | | 32–37 | 16 | 11.6 | — |
| Arc drop (volts) | | 48 | 20 | 12 | |
| Ionization potential (volts) | | 24.5 | 15.69 | 12.08 | 10.38 |
| Atomic wt. | | 4.002 | 39.94 | 131.3 | 200.6 |

* 125 ma.

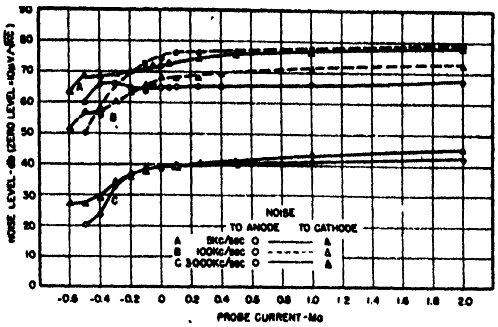
---

[6] Stuart Ballantine, Physics 4, 294 (1933).



Magnetic flux density =800 gauss at cathode
Cathode current =20 ma
Argon at 85 microns
Δ —noise measured between anode probe and cathode
O—noise measured between anode probe and anode

FIG. 4. Effect of probe current on noise at anode probe.



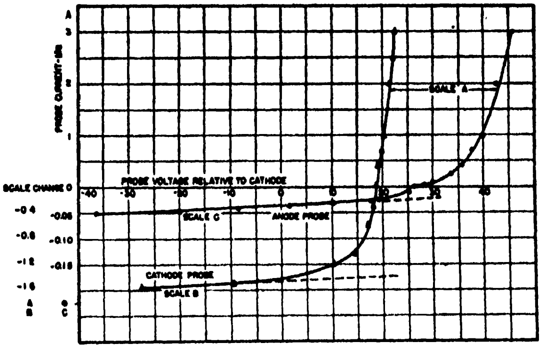
Magnetic flux density =800 gauss at cathode
Cathode current =20 ma
Argon at 85 microns
Δ —noise measured between cathode probe and cathode
O—noise measured between cathode probe and anode

FIG. 5. Effect of probe current on noise at cathode probe.

decrease and then increase the noise level over the entire spectrum. In no case where the magnetic field is equally strong at cathode and anode is the low frequency noise increased above the zero field value. On the other hand the high frequency noise can be increased markedly over the zero field value (Sections A and D, Fig. 2). The noise level in the high frequency region increases continuously up to the point where the magnetic field is strong enough to extinguish the discharge. The ultimate noise voltage level at 5 Mc/sec. may be 100 times the zero field level.

The noise level is also affected by gas pressure. With the discharge in a strong magnetic field, the noise increases continuously as the pressure is reduced, to the point where the pressure is too low for the discharge to be maintained in the presence of the field. Figure 3 shows a series of typical spectra obtained with an argon-filled experimental tube with a field of 860 gauss at the cathode. It will be noted that there is no "hump" in these spectra. This characteristic was noted for all tubes in which the electrode structure was such that the field was strong at one electrode only. In this case, the hump which is present at lower fields is obliterated by the high level of the low frequency noise.

The discharge current does not greatly effect the shape of the spectrum, if arc conditions are fully established. Both high and low frequency noise increase with increasing currents by a factor of not more than 3 when the arc current increases by a factor of 100. The noise is also substantially independent of the plasma volume, which was varied by a factor of 2000.



Magnetic flux density =800 gauss at cathode
Cathode current =20 ma
Argon at 85 microns

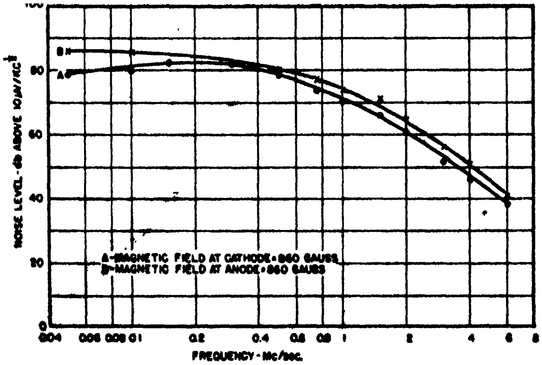
FIG. 6. Probe characteristics of experimental tube (Figs. 4 and 5).

The effect of different gases on the noise was studied. The general shape of the spectra is the same for all of the gases used, including helium, argon, xenon, and mercury. The high frequency noise (above 1 Mc/sec.) decreased as the atomic weight increased. No definite trend was noted in the low frequency noise. The results of these studies are summarized in Table I. The peak-to-peak voltage was measured on a wide-band oscilloscope. The peak-to-peak voltage is of the order of 85 percent of the arc drop. The arc drop itself is roughly proportional to the ionization potential of the gas used.

It was found that the amount of noise voltage measured between either cathode or anode, and an auxiliary probe electrode was highly dependent on the current collected by the probe. For these measurements, the main discharge was maintained between cathode and anode. The cathode-anode current and noise voltage were unaffected by the variations in the probe current. Two probes were used—one near the cathode and the other near the anode. The magnetic field was applied at the cathode and was perpendicular to the plane of the cathode probe. For each probe the noise was measured between probe and anode and probe and cathode. Figures 4 and 5 show the noise level at three frequencies as a function of probe current. The volt-ampere characteristics of the probes are shown in Fig. 6.

In all cases the noise was very low when the probe was made sufficiently negative that only positive-ion current was drawn. In general, the noise increases only when the probe begins to draw electron current. The noise at the cathode probe changes very gradually with current. The noise at the anode probe changes very abruptly when appreciable electron current is drawn. With magnetic field, the noise reaches its greatest value when the probe potential is such as to repel all but the fastest electrons. For these probe currents, conditions are such that very little noise is developed in the space charge region adjacent to the probe, since without the magnetic field, it has been shown that the maximum noise at the probe is developed only when its voltage is such as to accelerate electrons.[4]

Since the plasma acts as a "virtual" cathode with respect to the anode, it seemed probable that placing the magnetic field at the anode should produce the same effect on the noise spectrum as a field at the cathode. The experiment was tried, and the results were as expected (Fig. 7). In fact, the noise with the field at the anode was actually of higher level than that produced by a field at the cathode. Placing the magnetic field at any other position in the discharge decreased the over-all noise level, *i.e.*, the magnet had to be placed at a source of electrons to produce the high level noise. It has been observed elsewhere that the light from the discharge fluctuates with a "light" noise spectrum identical in shape with that of the electrical noise spectrum.[7] The light fluctuations were studied with a photo-cell. This effect was also noted in our investigation.



Argon at 85 microns
Anode current = 20 ma

FIG. 7. Noise spectrum of experimental tube for magnetic field at cathode (A) and at anode (B).

With a sufficiently strong magnetic field at the cathode, there is a narrow region of intense illumination adjacent to the cathode. Numerous experiments with fixed and moving probes show that the positive-ion density is abnormally high in the region close to the cathode when the magnetic field is applied. The positive-ion density was measured by the method of Langmuir and Mott-Smith.[8] For example, in one tube the positive-ion density near the cathode changed from $0.6\times10^{10}$ ions/cm$^3$ to $3.0\times10^{10}$ ions/cm$^3$ upon application of the magnetic field.

## DISCUSSION

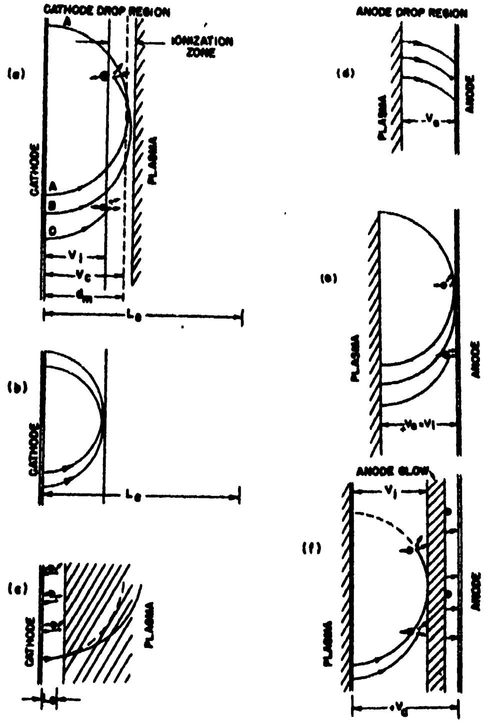
It is apparent that the generation of noise in the presence of the magnetic field is closely related to the space charge conditions at the cathode, or in particular, to the increased positive-ion density mentioned above. That the magnetic field must increase the positive-ion density in a narrow region near the cathode can be seen by considering Fig. 8. This figure presents a picture of the processes at a plane cathode and anode where each is in a strong magnetic field. The trajectory of an electron which does not collide with any atoms (A of Fig. 8a) will return to the cathode. The path will be determined by the electric and magnetic fields. In a uniform magnetic field, with plane electrodes and with no space charge or initial velocity, the path is cycloidal. If the electric field is distorted by

---

[7] Observed by Dr. R. W. Engstrom, Radio Corporation of America, Lancaster, Pennsylvania.

[8] I. Langmuir and H. Mott-Smith, Gen. Elec. Rev. 27, 449 (1924).

space charge, a direct calculation of the path is not possible unless a relation between electric field and distance is known. Thomson has shown that the paths in the dark space of a glow discharge in a transverse magnetic field are similar in form to cycloids.[9] It is probable that the paths do not differ greatly in the hot-cathode arc. If the pressure is such that the electrons make most of their collisions when near the point of maximum distance from the cathode, it is apparent that an intensely ionized zone will be found at this distance. If there is a very small spread in the initial velocities of the electrons, this zone will be very narrow. At higher pressures, the electrons will ionize much closer to the cathode, and the magnetic field will have little effect. Also, since the electric field is greater, the projected path will extend farther from the cathode (Fig. 8c). As the pressure is reduced, the point of maximum ionization moves out along the trajectory. At the same time, the maximum in the trajectory moves nearer the cathode because of the smaller electric field. At very low pressures the ionization must be occurring in a very narrow region, since reduction of the pressure makes it more and more probable that a collision will not occur until the electron has passed the maximum in the trajectory. In this portion of the path the electron is losing energy to the field. Thus, the electron must ionize before it reaches the point where its energy is less than the ionization potential (B of Fig. 8a). In the limiting case, immediately preceding extinction of the discharge, the zone of ionization must be very narrow. This picture assumes that most of the ionization results from single collisions. Multiple ionization, although certainly occurring, is not favored under conditions of low pressures. Such cumulative ionization as does occur would broaden the ionization zone.



(a) Cathode region, pressure = few hundred microns
A = trajectory of electron making no collisions
B = trajectory of electron making collision as it leaves ionization zone
C = trajectory of electron making collision as it enters ionization zone
(b) Cathode region, same pressure as (a), magnetic flux density = cut off value
(c) Cathode region, high pressure, same magnetic flux density as in (a)
Dotted trajectory same as A of (a)
Solid trajectory at high pressure
(d), (e), (f) Anode region, same pressure and magnetic flux density as in (a)
$L_e$ = electron mean free path
$V_i$ = ionization potential of gas
$V_c$ = cathode drop of potential
$d_m$ = mean distance of ionization zone from cathode
$V_a$ = anode drop of potential

FIG. 8. Details of electron paths with transverse magnetic field.

It is also possible for the magnetic field to establish a narrow zone of ionization at the anode. Possible space charge conditions are shown in Fig. 8 (d, e, f). If the anode drop is negative, electrons coming from the plasma are retarded, and little ionization occurs near the anode (Fig. 8d). If the anode drop is positive and slightly greater than the ionization potential, space charge conditions are similar to those at the cathode, since the plasma acts as a source of electrons (Fig. 8e). If the anode is small, a second ionized zone may be formed adjacent to it (Fig. 8f).

Dr. Felix Bloch suggested during the course of this research that the source of the increased noise is in the region of high space-charge density, where the ions in each element of volume can oscillate with a frequency determined by the local ion density, as in the plasma oscillations studied by Tonks and Langmuir. Since the ion density varies greatly over this region, a wide range of frequencies would be present instead

---

[9] J. J. Thomson, *Conduction of Electricity through Gases* (Cambridge University Press, London, 1933), third edition, Vol. II, p. 333 ff.

of the single characteristic frequency of the plasma. Tonks and Langmuir have shown that the plasma type oscillation can take place if the mean distance between ions is less than the Debye distance. For the ion densities and temperatures observed here, the Debye distance is several orders of magnitude greater than the interionic spacing. An unpublished analysis by Drs. Felix Bloch and Robert Weinstock[10] indicates that the highly ionized region may be as thin as $5\times10^{-5}$ cm and is capable of generating frequencies as high as 250 Mc/sec.

Additional evidence for the effect of the narrow ionization zone in producing the noise may be found in some additional experiments with oxide-coated cathodes. A number of commercial tubes were found deficient in high frequency noise unless the cathodes were properly treated to insure good emission. As might be expected, the arc drop was higher than for a normal tube at the same current. The higher arc drop is known to result in sputtering of the cathode materials which have much lower ionization potentials than argon. (Ionization potential of barium = 5.19 volts.) These sputtered atoms can be ionized over a broad region beginning close to the cathode. Thus, if there is an appreciable density of sputtered atoms, there will be a marked broadening of the ionization zone, with consequent lowering of the ion-density peak. This principle of the variation of high frequency noise level with cathode activation was used in designing an instrument for production testing of commercial tubes.

ACKNOWLEDGMENT

The authors wish to acknowledge the generous assistance of Dr. J. H. Van Vleck and Dr. Felix Bloch for many consultations during the course of this research. Acknowledgment is also made to Dr. Martin Freundlich and Mr. Stanley Ruthberg who built most of the experimental tubes used for these studies.

---

[10] Radio Research Lab. Rep., No. 411-232.

---

# Numerical Solution of Axisymmetrical Problems, with Applications to Electrostatics and Torsion

GEORGE SHORTLEY, *Mendenhall Laboratory of Physics, Ohio State University, Columbus, Ohio*

ROYAL WELLER, *Naval Ordnance Laboratory, Washington, D. C.*

PAUL DARBY,* *Mendenhall Laboratory of Physics, Ohio State University, Columbus, Ohio*

AND

EDWARD H. GAMBLE,** *Mendenhall Laboratory of Physics, Ohio State University, Columbus, Ohio*

(Received September 19, 1946)

Numerical methods are given for solution of axisymmetrical problems involving the partial differential equation

$$\frac{\partial^2\psi}{\partial z^2}+\frac{\partial^2\psi}{\partial\rho^2}+\frac{K}{\rho}\frac{\partial\psi}{\partial\rho}=0,$$

where $\rho$ is the radial coordinate and $z$ the coordinate parallel to the axis. The various values of $K$ which occur in physical situations are discussed, and common iteration methods for handling these problems are given. For Laplace's equation, $K=1$. For the Stoke's stream function, $K=-1$, but it is pointed out that for numerical work a new function, called the *flow-disturbance function*, having $K=3$, is more tractable. Similarly a new function with $K=5$, the *stress-concentration function*, is much easier to compute than the usual stress function ($K=-3$) for the case of the torsion of a circular shaft of varying diameter. The methods are illustrated by computation of the equipotentials for an electron lens, and by a complete computation of the stresses and strains in a particular grooved circular shaft under torsion.

---

* Now at Remington Arms Company, Bridgeport, Connecticut.

** Now at Bell Telephone Laboratories, Murray Hill, New Jersey.

## 1. INTRODUCTION

THIS paper represents an extension of the numerical methods for handling partial differential equations in two dimensions, which were developed in recent work on "Numerical Solution of Laplace's and Poisson's Equations, with Applications to Photoelasticity and Torsion."[1] For the basis of some of the methods used in the present paper, reference should be made to this previous work. On the other hand, the method of extrapolation developed in Section 6 of the present paper is an improvement over that previously given, and can be used in the solution of Laplace's and Poisson's equations in two dimensions as well as in the cases discussed here.

An axisymmetrical problem is essentially two-dimensional, since the functions which arise depend only on the coordinate $z$, measured parallel to the axis, and the radius $\rho$, measured normal to the axis, and are independent of the azimuthal angle $\varphi$. The differential equations which govern many functions of technical interest are of the form

$$\frac{\partial^2\psi}{\partial z^2}+\frac{\partial^2\psi}{\partial \rho^2}+\frac{K}{\rho}\frac{\partial\psi}{\partial\rho}=0, \qquad (1)$$

which we abbreviate to

$$\Delta'\psi+(K/\rho)\psi_\rho=0. \qquad (1')$$

Here the constant $K$ takes on the following values:

$K=1$, Laplace's equation, applicable to the potentials of axially symmetric electrostatic, magnetostatic, and gravitational fields; and the steady flow of heat, electric current, and ideal fluid. This case is discussed in Section 7, with an application to an electron-lens system.

$K=-1$, Stoke's stream function, which is constant along a stream line in the steady flow of heat, electric current, and ideal fluid. This function, while it satisfies simpler boundary conditions than the potential, has the disadvantage for numerical work of rapid variation with $\rho$; for a unidirectional stream it varies as $\rho^2$. Hence it is preferable to introduce a new function, the *flow-disturbance function*, which is the stream function divided by $\rho^2$. This function is constant in regions free from disturbance and satisfies the above equation with

$K=3$; the use of the flow-disturbance function for numerical work is exactly analogous to the use of the stress-concentration function which is discussed in detail in this paper.

$K=-3$, a function which governs the stresses in a circular shaft of varying diameter. We introduce a new function, the *stress-concentration function*, more suitable for numerical work, which satisfies the equation with

$K=5$; the use of the stress-concentration function is the subject of Sections 8–10 of this paper.

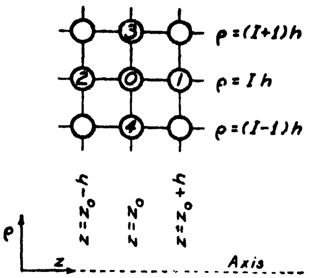


FIG. 1. Notation for an ordinary point of the net.

Because these problems differ only in the value of $K$ in (1), the methods of numerical solution will be similar and can be discussed without specifying the value of $K$.

It is well to point out at this time that any numerical integration of an elliptic partial differential equation in two or more dimensions is bound to be very tedious and time-consuming if high accuracy is desired simply because of the number of net points which must be used. The more irregular the boundaries, the more points which must be used if accuracy is to be obtained in the vicinity of the irregularities. But it seems important to know that a computation can be carried through numerically when required, and to investigate the most expeditious manner of making the computation. The procedures given below and in reference 1 are believed to contain most of the formal tricks for speeding up the computations which have been discovered to date and to constitute the best available computational scheme. It is hoped that some day high speed electronic computing equipment will be developed with sufficient internal-memory capacity to handle data on a large net; in this case it may be advisable to ignore tricks, use the simplest possible improvement formula at each point, and let the machine work somewhat

---

[1] G. Shortley and R. Weller, J. App. Phys. 9, 334 (1938); J. App. Mech. 6, A-71 (1939); Weller, Shortley, and Fried, J. App. Phys. 11, 283 (1940). These three papers have been combined in the Ohio State University Engineering Experiment Station Bulletin No. 107 (1940); revised, 1942. In these papers will be found complete references to earlier work.

longer, or to use some non-iterative procedure such as a direct numerical solution of the difference equations.

## 2. THE ITERATION PROCEDURE

We first replace the continuum of points $(z, \rho)$ by the discrete set of points at the intersections of a net of intervals $h$:

$$z=0,\ \pm h,\ \pm 2h,\ \pm 3h,\ \cdots,$$
$$\rho=0,\ h,\ 2h,\ \cdots,\ Ih,\ \cdots.$$

The differential equation is then replaced by a difference equation which relates the value of $\psi$ at one point to its values at neighboring points. The solution of the difference equation is then an approximation to the solution of the differential equation—the finer the net, the better the approximation.

For example, consider the five numbered points in the diagram (Fig. 1). Denote the function values at these points by $\psi_0$, $\psi_1$, $\psi_2$, $\psi_3$, and $\psi_4$. At the point 0 $(z=z_0,\ \rho=Ih)$, Eq. (1′) may be written as

$$\Delta'\psi_0+(K/Ih)\psi_{\rho 0}=0$$

or

$$h^2\Delta'\psi_0+(K/I)h\psi_{\rho 0}=0. \tag{2}$$

Now on any quadric surface passing through the five function values $\psi_0, \cdots, \psi_4$, the following relations are satisfied:

$$h^2\Delta'\psi_0=-4\psi_0+\psi_1+\psi_2+\psi_3+\psi_4,$$
$$h\psi_{\rho 0}=\tfrac{1}{2}\psi_3-\tfrac{1}{2}\psi_4.$$

Hence the differential equation (2) is approximated—to the accuracy with which the function $\psi$ is representable by a quadric in the region containing these five points—by the difference equation

$$-4\psi_0+\psi_1+\psi_2+\psi_3+\psi_4+(K/I)(\tfrac{1}{2}\psi_3-\tfrac{1}{2}\psi_4)=0 \tag{3}$$

or

$$8I\psi_0=2I\psi_1+2I\psi_2+(2I+K)\psi_3+(2I-K)\psi_4. \tag{4}$$

This difference equation holds at all ordinary points of the net, but needs modifications, which will be considered later, at points near the axis or near a material boundary.

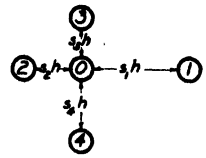
The iteration procedure, originated by Liebmann,[1] for finding the function values which satisfy this difference equation at all net points, proceeds as follows: Start with any arbitrary "trial" function $\psi$—a guess. Then go over the points of the net, in some definite order, replacing the $\psi$ value at each net point by a new value calculated from (4) in terms of the values at neighboring points. This will give an "improved" function $\psi'$. In calculating the improved value at a given point, the previously altered values are used whenever they are available. This procedure is then iterated. Successive traverses of the net give successively functions $\psi', \psi'', \psi''', \cdots$ which approach the true solution of the difference equation in a manner to be considered in Section 6.

We shall now consider modifications of the difference equation (3), necessary at points near the axis or near a boundary, and modifications leading to more rapid convergence. The essence of the iteration procedure is that at any point the difference equation, solved for $\psi_0$ as in (4), may be used as an "improvement formula." A proof of convergence, by this process, to the exact solution of the difference equation, and a method of investigating the rate of convergence, are developed in reference 1. This theory assumes that the improvement formula expresses the value at a point as a *true weighted average* of the neighboring values, *with all weights positive.* We note that the coefficient of $\psi_0$, in (4) and the following improvement formulas, equals the sum of the coefficients of the neighboring values $\psi_1$, $\psi_2$, $\cdots$. A little experimentation shows that if negative coefficients are permitted, the procedure may not converge, so we reject any improvement formula which involves negative coefficients and attempt to replace it by an alternative formula in which all the coefficients are positive.

## 3. IMPROVEMENT FORMULAS

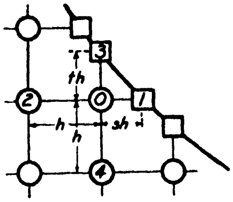
The following formulas are derived by passing a quadric through $\psi_0, \psi_1, \psi_2, \cdots$ which is required to satisfy the differential equation at point 0. The mechanism for accomplishing this is discussed in Appendix 1. Certain of the coefficients in these formulas, as well as in (4), become negative for high $K$ and low $I$ values. Formulas with negative coefficients may be used for interpolation but not for improvement. Modifications necessary to eliminate these negative coefficients will be considered later for each $K$ value separately. *In each diagram below, the coordinate axes are assumed to run as in Fig. 1, with the symmetry axis horizontal and below the diagram, and with coordinates $z=z_0$, $\rho=Ih$, for the point 0.*

*Unequal spacing:*



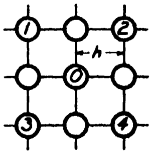
$$\left(\frac{2I}{s_1s_2}+\frac{2I+K(s_4-s_3)}{s_3s_4}\right)\psi_0=\frac{2I}{s_1(s_1+s_2)}\psi_1+\frac{2I}{s_2(s_1+s_2)}\psi_2+\frac{2I+Ks_4}{s_3(s_3+s_4)}\psi_3+\frac{2I-Ks_3}{s_4(s_3+s_4)}\psi_4. \tag{5}$$

*Point near boundary:*



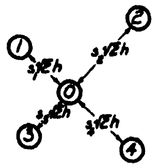
$$\left(\frac{2I}{s}+\frac{2I+K(1-t)}{t}\right)\psi_0=\frac{2I}{s(1+s)}\psi_1+\frac{2I}{1+s}\psi_2+\frac{2I+K}{t(1+t)}\psi_3+\frac{2I-Kt}{1+t}\psi_4. \tag{6}$$

*Diagonal formula:*



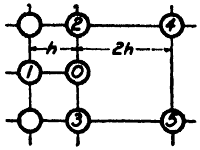
$$8I\psi_0=(2I+K)\psi_1+(2I+K)\psi_2+(2I-K)\psi_3+(2I-K)\psi_4. \tag{7}$$

*Diagonal formula, unequal spacing:*

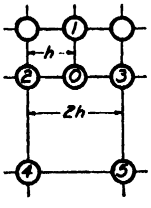


$$\left(\frac{2I+K(s_4-s_1)}{s_1s_4}+\frac{2I+K(s_3-s_2)}{s_3s_2}\right)\psi_0=\frac{2I+Ks_4}{s_1(s_1+s_4)}\psi_1+\frac{2I+Ks_3}{s_2(s_2+s_3)}\psi_2+\frac{2I-Ks_2}{s_3(s_2+s_3)}\psi_3+\frac{2I-Ks_1}{s_4(s_1+s_4)}\psi_4. \tag{8}$$

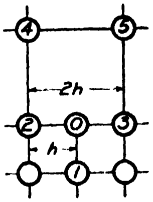
*Change of net spacing:*



$$16I\psi_0=4I\psi_1+(5I+3K)\psi_2+(5I-3K)\psi_3+I\psi_4+I\psi_5; \tag{9}$$

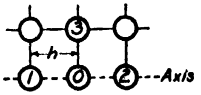


$$(32I+8K)\psi_0=(8I+8K)\psi_1+(10I+K)\psi_2+(10I+K)\psi_3+(2I-K)\psi_4+(2I-K)\psi_5; \tag{10}$$

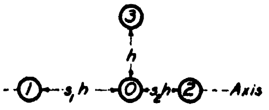


$$(32I-8K)\psi_0=(8I-8K)\psi_1+(10I-K)\psi_2+(10I-K)\psi_3+(2I+K)\psi_4+(2I+K)\psi_5. \tag{11}$$

*Axis points* (obtained from a quadric, even in $\rho$, which satisfies the differential equation everywhere):

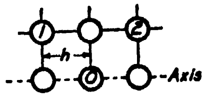


$$(2K+4)\psi_0=\psi_1+\psi_2+(2K+2)\psi_3; \tag{12}$$



$$\left(1+K+\frac{1}{s_1s_2}\right)\psi_0=\frac{1}{s_1(s_1+s_2)}\psi_1+\frac{1}{s_2(s_1+s_2)}\psi_2+(1+K)\psi_3. \tag{13}$$

*Axis point, diagonal* (obtained from the linear solution $\psi=\psi_0+Az$; not to be used except for interpolation):



$$\psi_0=\tfrac{1}{2}(\psi_1+\psi_2). \tag{14}$$

## 4. NINE-BLOCK FORMULAS

As explained in reference 1, the convergerce of the iteration procedure is much more rapid if we improve, not one point at a time, but a block of points all at once. A block of nine points is the maximum that can be handled conveniently. In the case of Laplace's equation, the use of nine-blocks was found to cut the labor by a factor of 3.5, and the saving of time is of the same order of magnitude for these axially symmetric problems.

The idea is to find improved values for all the nine circled points of Fig. 2 in terms of the values at the surrounding points, marked by squares. We start by writing the difference equation (4) for each of the nine points in terms of its neighbors. Elimination of the $\psi$ values at points $b$, $B$, $c$, $d$, $D$ then leaves a formula for $\psi_0$ in terms of only the outside points $e$, $E$, $f$, $F$, $g$, $G$, $m$. This formula gives an improved value for $\psi_0$ in which the difference equation is satisfied simultaneously at all nine points of the block. The coefficients in this improvement formula are fifth-order polynomials in $I$ and $K$ (see Appendix 2), so we give in Tables I and II the numerical values needed for $K=1$ and $K=5$. These tables give the coefficients by which each of the 12 boundary points are to be multiplied to get the improved center-point value. The coefficients given for points $F$, $G$, $m$, $g$, and $f$ are each to be used twice.

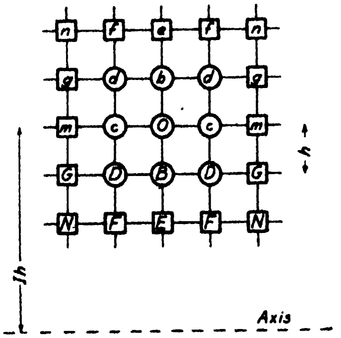


FIG. 2. The nine-block.

TABLE I. Coefficients for improvement of nine-block center point ($K=1$).

| | Coefficients of boundary points in Fig. 2 | | | | | | |
|---|---|---|---|---|---|---|---|
| *I* | *E* | *F* | *G* | *m* | *g* | *f* | *e* |
| 0 | -- | — | — | 0.042553 | 0.106383 | 0.159574 | 0.382979 |
| 2 | 0.048393 | 0.024451 | 0.048902 | 0.131764 | 0.081503 | 0.095087 | 0.188193 |
| 3 | .079053 | .039681 | .052908 | .127476 | .074072 | .083331 | .166010 |
| 4 | .091719 | .045955 | .055147 | .126312 | .070903 | .077993 | .155661 |
| 5 | .098824 | .049476 | .056544 | .125818 | .069110 | .074869 | .149543 |
| 6 | .103402 | .051747 | .057497 | .125560 | .067951 | .072805 | .145479 |
| 7 | .106607 | .053338 | .058187 | .125408 | .067139 | .071335 | .142578 |
| 8 | .108979 | .054517 | .058710 | .125311 | .066538 | .070235 | .140400 |
| 9 | .110807 | .055425 | .059120 | .125244 | .066075 | .069379 | .138704 |
| 10 | .112259 | .056147 | .059450 | .125198 | .065708 | .068695 | .137346 |
| 11 | .113441 | .056735 | .059721 | .125163 | .065409 | .068134 | .136233 |
| 12 | .114422 | .057223 | .059948 | .125137 | .065161 | .067667 | .135305 |
| 13 | .115249 | .057635 | .060141 | .125116 | .064952 | .067272 | .134519 |
| 14 | .115955 | .057987 | .060307 | .125100 | .064774 | .066933 | .133844 |
| 15 | .116567 | .058291 | .060450 | .125087 | .064619 | .066639 | .133259 |
| 16 | .117100 | .058557 | .060577 | .125077 | .064485 | .066381 | .132746 |
| 17 | .117570 | .058792 | .060688 | .125068 | .064366 | .066154 | .132294 |
| 18 | .117988 | .059000 | .060787 | .125060 | .064261 | .065952 | .131891 |
| 19 | .118361 | .059185 | .060876 | .125054 | .064167 | .065771 | .131531 |
| 20 | .118696 | .059352 | .060957 | .125049 | .064083 | .065608 | .131206 |
| 21 | .118999 | .059504 | .061029 | .125044 | .064006 | .065461 | .130913 |
| 22 | .119274 | .059641 | .061095 | .125040 | .063937 | .065327 | .130645 |
| 23 | .119525 | .059766 | .061156 | .125037 | .063874 | .065205 | .130401 |
| 24 | .119755 | .059881 | .061211 | .125034 | .063816 | .065092 | .130177 |
| 25 | .119966 | .059986 | .061262 | .125031 | .063763 | .064989 | .129971 |
| 26 | .120161 | .060083 | .061309 | .125029 | .063714 | .064894 | .129781 |
| 27 | .120341 | .060173 | .061353 | .125027 | .063668 | .064805 | .129605 |
| 28 | .120509 | .060257 | .061394 | .125025 | .063626 | .064723 | .129441 |
| 29 | .120665 | .060335 | .061432 | .125023 | .063587 | .064647 | .129289 |
| 30 | .120810 | .060407 | .061467 | .125022 | .063551 | .064576 | .129147 |

TABLE II. Coefficients for improvement of nine-block center point ($K=5$).

| | Coefficients of boundary points in Fig. 2 | | | | | | |
|---|---|---|---|---|---|---|---|
| *I* | *E* | *F* | *G* | *m* | *g* | *f* | *e* |
| 0 | - | -- | — | 0.043478 | 0 | 0.130435 | 0.652174 |
| 4 | 0.007149 | 0.003466 | 0.020797 | 0.107452 | 0.090121 | 0.135182 | 0.278813 |
| 5 | .022148 | .010862 | .028965 | .113597 | .086895 | .123101 | .251011 |
| 6 | .035056 | .017296 | .034593 | .117015 | .084011 | .114015 | .231084 |
| 7 | .045545 | .022552 | .038661 | .119104 | .081618 | .107123 | .216339 |
| 8 | .054042 | .026822 | .041722 | .120471 | .079652 | .101777 | .205069 |
| 9 | .060996 | .030320 | .044102 | .121414 | .078027 | .097534 | .196210 |
| 10 | .066762 | .033223 | .046002 | .122090 | .076669 | .094094 | .189080 |
| 11 | .071605 | .035663 | .047551 | .122592 | .075522 | .091256 | .183225 |
| 12 | .075724 | .037738 | .048838 | .122975 | .074542 | .088877 | .178336 |
| 13 | .079265 | .039522 | .049923 | .123273 | .073696 | .086856 | .174196 |
| 14 | .082339 | .041071 | .050850 | .123510 | .072958 | .085118 | .170645 |
| 15 | .085032 | .042427 | .051651 | .123702 | .072311 | .083609 | .167569 |
| 16 | .087409 | .043624 | .052349 | .123859 | .071738 | .082287 | .164877 |
| 17 | .089522 | .044688 | .052964 | .123989 | .071227 | .081120 | .162504 |
| 18 | .091412 | .045640 | .053509 | .124098 | .070769 | .080081 | .160395 |
| 19 | .093112 | .046496 | .053995 | .124190 | .070357 | .079152 | .158509 |
| 20 | .094650 | .047270 | .034432 | .124269 | .069983 | .078315 | .156814 |
| 21 | .096047 | .047973 | .054826 | .124337 | .069644 | .077558 | .155280 |
| 22 | .097322 | .043614 | .055183 | .124396 | .069333 | .076869 | .153888 |
| 23 | .098490 | .049201 | .055509 | .124447 | .069048 | .076241 | .152617 |
| 24 | .099564 | .049741 | .055808 | .124492 | .068786 | .075665 | .151453 |
| 25 | .100555 | .050240 | .056081 | .124532 | .068544 | .075135 | .150382 |
| 26 | .101471 | .050700 | .056334 | .124567 | .068320 | .074646 | .149395 |
| 27 | .102322 | .051128 | .056567 | .124599 | .068112 | .074193 | .148481 |
| 28 | .103113 | .051526 | .056784 | .124627 | .067918 | .073773 | .147634 |
| 29 | .103852 | .051897 | .056985 | .124652 | .067737 | .073381 | .146845 |
| 30 | .104542 | .052244 | .057172 | .124675 | .067567 | .073016 | .146109 |

After the center value $\psi_0$ has been determined in this way, the four corner values $d$, $D$, are determined from it and the outside values at $e$, $E$, $m$, $n$, $N$ by using the diagonal formula (7); finally the side points $b$, $B$, $c$ are obtained from the ordinary improvement formula (4). In this way improved values for all nine points of the block are filled in without using any of the old values at these points.

## 5. USE OF THE DIFFERENCE FUNCTION

A great saving in labor is achieved by carrying out operations, not on the function $\psi$ itself, but on the "difference function," which represents the change in $\psi$ with one traverse. We suppose that the net has been completely laid out, the improvement formula for each point and a definite order of traversing the points and blocks decided on, and the initial trial function chosen. Suppose that the net were traversed several times, to obtain successively the improved functions $\psi'$, $\psi''$, $\cdots$. We call the change in function value at each point, $\delta=\psi'-\psi$, the *difference function*. Let $\delta'$ be the next change or difference: $\delta'=\psi''-\psi'$. Then if, after one traverse, we compute $\delta$ at each point by subtraction, we can compute what the next change $\delta'$ will be by applying our improvement formulas directly to $\delta$. For, application of the improvement process to $\psi$ gives $\psi'$, application to $\psi'$ gives $\psi''$, application to the difference $\delta=\psi'-\psi$ will give the difference $\delta'=\psi''-\psi'$.

Hence after *one* traverse, we can compute all successive changes which $\psi$ will undergo by operating on the difference function $\delta$ in place of operating on $\psi$ itself. There are a number of factors which make it much simpler to work with the difference function than with $\psi$. First, the difference function usually has fewer significant figures than has $\psi$. Second, the boundary values of the difference function are all zero; this gives many zero terms in the improvement formulas. Finally, we know the value to which the difference function will converge after a large number of improvements, *viz.*, to the value *zero* at each point, because of the zero boundary values. This convergence to zero takes place in a regular manner which permits convenient extrapolation.

If the successive difference functions are denoted by $\delta=\psi'-\psi$, $\delta'$, $\delta''$, $\delta'''$, $\cdots$, the final solution of the problem for this net will be obtained by the addition

$$\Psi=\psi'+\delta'+\delta''+\delta'''+\cdots. \qquad (15)$$

## 6. EXTRAPOLATION

As discussed in reference 1, it is possible to find certain functions with zero boundary values which, when improved, will be reduced to a constant fraction of themselves without being changed in shape. Such a function is called a characteristic function of the improvement process. The different characteristic functions are denoted by $\phi^{(1)}$, $\phi^{(2)}$, $\phi^{(3)}$, $\cdots$. On improvement these go over to

$$\phi^{(1)\prime}=\lambda_1\phi^{(1)}, \quad \phi^{(2)\prime}=\lambda_2\phi^{(2)}, \quad \text{etc.}$$

The $\lambda$ values (characteristic values) will all be less than unity, and one of them, say $\lambda_1$, will be the largest. It is now possible to make a sort of Fourier expansion of the actual difference function $\delta$ in terms of the characteristic functions,[2] i.e.,

$$\delta=a_1\phi^{(1)}+a_2\phi^{(2)}+a_3\phi^{(3)}+\cdots.$$

On improvement this becomes successively

$$\begin{aligned}\delta'&=a_1\lambda_1\phi^{(1)}+a_2\lambda_2\phi^{(2)}+a_3\lambda_3\phi^{(3)}+\cdots,\\ \delta''&=a_1\lambda_1^2\phi^{(1)}+a_2\lambda_2^2\phi^{(2)}+\cdots,\\ \delta'''&=a_1\lambda_1^3\phi^{(1)}+a_2\lambda_2^3\phi^{(2)}+\cdots, \quad \text{etc.}\end{aligned}$$

Thus after $k$ improvements, the coefficients of the various characteristic functions in $\delta$ are decreased by factors $\lambda^k$. After a time only the function with the highest characteristic value will remain, and $\delta^{[k]}$ will become a fairly smooth, pillow-shaped, one-signed function decreasing uniformly by a factor $\lambda_1$ at each iteration. At this stage one has

$$\begin{aligned}\delta^{[k]}&=c\phi^{(1)},\\ \delta^{[k+1]}&=\lambda_1c\phi^{(1)},\\ \delta^{[k+2]}&=\lambda_1^2c\phi^{(1)}, \quad \text{etc.}\end{aligned}$$

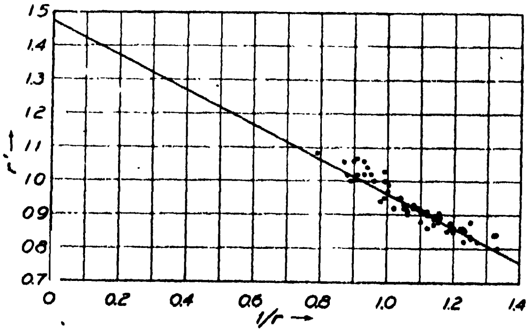


FIG. 3. Extrapolation. Plot of ratios in Eq. (18) for first extrapolation on net shown in Fig. 7. The straight line has the equation $r'=1.475-0.513(1/r)$ so $\lambda_1\approx0.914$, $\lambda_2\approx0.561$, and the right member of Eq. (19) becomes $-13\delta^{[k-1]}+26\delta^{[k]}$.

---

[2] This statement is not mathematically rigorous, but is adequate for the practical application of the extrapolation scheme. For discussion of this point, see the first paper of reference 1.

so that

$$\delta^{[k]}+\delta^{[k+1]}+\delta^{[k+2]}+\cdots$$
$$=c\phi^{(1)}(1+\lambda_1+\lambda_1^2+\cdots)=\delta^{[k]}/(1-\lambda_1),$$

which furnishes the evaluation of the infinite sum occuring in (15).

This was the method of extrapolation contemplated in reference 1. But is it not necessary to wait for all characteristic functions except $\phi^{(1)}$ to become small before extrapolating. For a fine net ($\lambda$ values close to unity), it will cut the labor by a large factor if we can extrapolate when *two* of the functions, $\phi^{(1)}$ and $\phi^{(2)}$, remain of importance. Let us suppose that we have improved the difference function $k$ times and that we may write

$$\begin{aligned}\delta^{[k-2]}&=c_1\phi^{(1)}+c_2\phi^{(2)},\\ \delta^{[k-1]}&=\lambda_1c_1\phi^{(1)}+\lambda_2c_2\phi^{(2)},\\ \delta^{[k]}&=\lambda_1^2c_1\phi^{(1)}+\lambda_2^2c_2\phi^{(2)},\quad\text{etc.}\end{aligned}\tag{16}$$

Let us define two successive ratios by

$$r=\delta^{[k-1]}/\delta^{[k-2]};\quad r'=\delta^{[k]}/\delta^{[k-1]}.\tag{17}$$

If $c_1\phi^{(1)}$ and $c_2\phi^{(2)}$ are eliminated from these two ratios, we find that if (16) are satisfied, then

$$r'=(\lambda_1+\lambda_2)-\lambda_1\lambda_2(1/r).\tag{18}$$

Hence we can test whether a net has reached the stage where the differences obey fairly well the relations (16) by computing $r$ and $r'$ at various points and plotting $r'$ against $(1/r)$. If (16) are exactly obeyed, these points will fall on a straight line of slope $(-\lambda_1\lambda_2)$ and intercept $(\lambda_1+\lambda_2)$. This plot will not only tell whether (16) are approximately valid, but will determine the values of $\lambda_1$ and $\lambda_2$ to use in the extrapolation. Such a plot is given in Fig. 3 by way of illustration.

With characteristic values determined as above for the net as a whole, the values of $c_1\phi^{(1)}$ and $c_2\phi^{(2)}$ at any point are fixed by any two successive $\delta$'s of (16). Algebraic manipulation shows that if (16) are obeyed, then

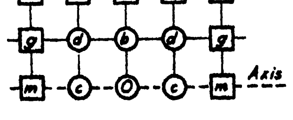


FIG. 4. A nine-block split by the axis.

$$\delta^{[k-1]}+\delta^{[k]}+\delta^{[k+1]}+\delta^{[k+2]}+\cdots$$
$$=\delta^{[k-1]}\frac{1-\lambda_1-\lambda_2}{(1-\lambda_1)(1-\lambda_2)}+\frac{\delta^{[k]}}{(1-\lambda_1)(1-\lambda_2)}.\tag{19}$$

Hence the extrapolation proceeds as follows: Suppose that we have computed

$$\begin{gathered}\psi,\ \psi',\\ \delta,\ \delta',\ \delta'',\ \cdots,\ \delta^{[k-2]},\ \delta^{[k-1]},\ \delta^{[k]},\end{gathered}\tag{20}$$

and believe that the differences are proceeding regularly enough to try an extrapolation under the assumption of (16). Compute $r$ and $r'$ for all points of the net which give enough significant figures to be useful in $\lambda$-value determinations. Make a plot like Fig. 3; from it determine $\lambda_1$ and $\lambda_2$. Compute the coefficients $(1-\lambda_1-\lambda_2)/(1-\lambda_1)(1-\lambda_2)$ and $1/(1-\lambda_1)(1-\lambda_2)$ in (19). Then compute the extrapolated function value $\Psi$ at each point by adding (see (15))

$$\Psi=\psi'+\delta'+\delta''+\cdots+\delta^{[k-2]}$$
$$+\frac{1-\lambda_1-\lambda_2}{(1-\lambda_1)(1-\lambda_2)}\delta^{[k-1]}+\frac{1}{(1-\lambda_1)(1-\lambda_2)}\delta^{[k]}.\tag{21}$$

Note that while the three last differences of (20) are assumed to obey (16) for the purpose of computing $\lambda_1$ and $\lambda_2$, only the last two differences $\delta^{[k-1]}$ and $\delta^{[k]}$ need obey (16) for the extrapolation (21) to be valid.

### 7. $K=1$, LAPLACE'S EQUATION

In this case we run into no difficulty with negative coefficients under any circumstance. Table I gives the coefficients by which the values at the outside points of Fig. 2 are to be multiplied to obtain the improved value $\psi_0$ at the center point of the block. If these coefficients are used as multipliers on the computing machine, a check of the computation is furnished by the fact that their sum, after all 12 values have been used, is unity. In addition to these values, it is convenient to have those for a block split by the axis as in Fig. 4. For this block the center point

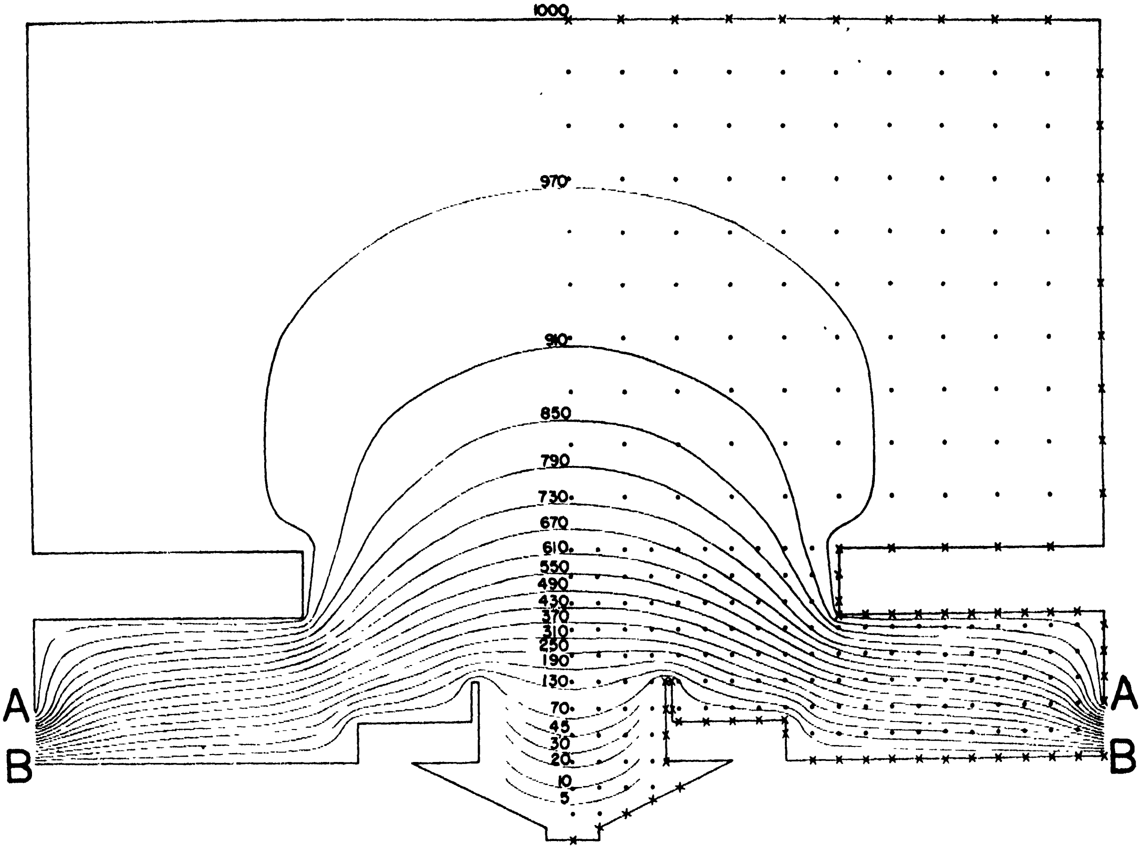


FIG. 5. Equipotential surfaces for an electron lens.

is given by

$$376\psi_0 = 16\psi_m + 40\psi_g + 60\psi_f + 144\psi_e + 60\psi_f + 40\psi_g + 16\psi_m, \qquad (22)$$

where the boundary points are taken in order around the block. These coefficients, divided by 376, are listed for $I=0$ in Table I.

As an example of the application of this numerical method we print Fig. 5, which shows the equipotential surfaces for an electron lens in whose properties we were interested. The axis in this figure is vertical. The potential of the upper electrode $AA$ is taken as 1000 volts, that of the lower electrode $BB$ as zero. The net points and boundary points used in the computation are indicated by circles and crosses, respectively. The equipotential surfaces are accurate for the difference equation on these net points; the departure from the solution of the differential equation is probably about one percent over most of the net. We were particularly interested in the degree of penetration of the field into the almost enclosed spaces in the two electrodes.

## 8. THE TORSION OF AN AXIALLY SYMMETRIC SHAFT

The stresses in a circular shaft of variable diameter subjected to torsion may be expressed[3] in terms of a function $\phi(\rho, z)$ which satisfies the differential equation

$$\frac{\partial^2\phi}{\partial\rho^2} + \frac{\partial^2\phi}{\partial z^2} - \frac{3}{\rho}\frac{\partial\phi}{\partial\rho} = 0. \qquad (23)$$

In terms of this function, the stresses have the values:

Normal stresses: $\sigma_\rho = \sigma_z = \sigma_\varphi = 0.$

---

[3] S. Timoshenko, *Theory of Elasticity* (McGraw-Hill Book Company, Inc., New York, 1934), p. 276 *et seq.*; our definition of $\phi$ differs by a factor $2\pi$ from that of Timoshenko. Physically, $\phi(\rho, z)$ is the torque transmitted across a surface bounded by the circle $\rho = \rho$, $z = z$.

Shears: $\tau_{\rho z}=0,$

$$\tau_{\varphi\rho}=-\frac{1}{2\pi\rho^2}\frac{\partial\phi}{\partial z}, \tag{24}$$

$$\tau_{\varphi z}=\frac{1}{2\pi\rho^2}\frac{\partial\phi}{\partial\rho}$$

Boundary conditions require that $\phi=0$ along the axis or on the inner boundary if the shaft is partially or entirely hollow; $\phi=M$, the torque, on the outer boundary. In a diametral $(\rho, z)$ plane, lines of constant $\phi$ are in the direction of the maximum shear.

For a straight circular shaft of radius $a$, the function $\phi$ is $M\rho^4/a^4$. The only component of stress is $\tau_{\varphi z}=2M\rho/\pi a^4$.

Since $\tau_{\varphi z}$ must vanish at $\rho=0$, we see from (24) that for any shaft $\phi$ must vary like $\rho^4$ near the axis. A function varying this rapidly is inconvenient to handle by numerical methods; it is simpler to work with the function

$$\Psi=\phi/\rho^4 \tag{25}$$

which will start like $A+B\rho^2+\cdots$ (where $A$ and $B$ are functions of $z$) near the axis.

The function $\Psi$ will satisfy the differential equation

$$\frac{\partial^2\Psi}{\partial\rho^2}+\frac{\partial^2\Psi}{\partial z^2}+\frac{5}{\rho}\frac{\partial\Psi}{\partial\rho}=0. \tag{26}$$

Boundary values for $\Psi$ are determined from those of $\phi$ according to (25); along the axis $\Psi$ does not vanish but $\partial\Psi/\partial\rho=0$, and in fact, if $\Psi$ be expanded in a power series, only even powers of $\rho$ may occur.

For a straight circular shaft,

$$\Psi=\text{const.}=M/a^4.$$

For a grooved or filleted shaft with long straight cylindrical end sections, $\Psi$ is constant except in the vicinity of the groove or fillet where the stresses depart from the circular-shaft values. For this reason, we call $\Psi$ the *stress-concentration function.*

To complete this discussion of elastic theory, we note that if $\Theta(\rho, z)$ represents the angle of twist in radians, at any point, this function is related to the stress function $\phi(\rho, z)$ as follows:[8]

$$\frac{\partial\Theta}{\partial\rho}=-\frac{1}{2\pi G\rho^3}\frac{\partial\phi}{\partial z},$$

$$\frac{\partial\Theta}{\partial z}=\frac{1}{2\pi G\rho^3}\frac{\partial\phi}{\partial\rho},$$

where $G$ is the modulus of elasticity in shear. Hence, for a straight circular shaft of radius $a$, $\Theta$ is $2Mz/\pi a^4G$. From the above relations, it follows that in the $\rho$, $z$ plane, lines of constant $\Theta$ are orthogonal to lines of constant $\phi$.

### 9. $K=5$, THE STRESS-CONCENTRATION FUNCTION

We consider next the application of the numerical methods of Sections 2–6 to the solution of Eq. (26).

Coefficients in the simple improvement formulas (4) and (7) become negative at $I=1$ and $I=2$; special formulas for these points are therefore needed.

For $I=2$, formula (4) gives (cf. Fig. 1)

$$16\psi_0=4\psi_1+4\psi_2+9\psi_3-\psi_4. \tag{27}$$

This formula is not adapted to the iteration process because of the negative coefficient. It was derived by passing a quadric function through points (0), (1), (2), (3), and (4) which satisfied the differential equation at (0). An alternative formula (see Appendix 1) is obtained by passing a fourth-order polynomial *solution* of the differential equation through the five points. For such a solution,

$$4\psi_0=10\psi_1+10\psi_2+9\psi_3-25\psi_4. \tag{28}$$

Use of such fourth-order functions is not advisable in numerical work except on extremely fine nets. But the following linear combination of (27) and (28) eliminates the negative coefficient of $\psi_4$, and involves much more of (27) than (28):

$$22\psi_0=5\psi_1+5\psi_2+12\psi_3 \quad (K=5,\ I=2). \tag{29}$$

It is proposed to use this as the ordinary improvement formula at $I=2$. Similar considerations will not give a diagonal formula for $I=2$ which is free from negative coefficients. Hence,

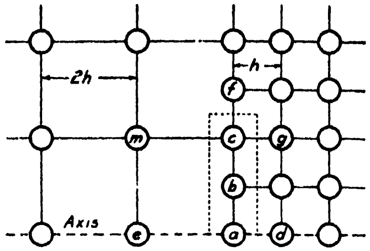


FIG. 6. Change of net spacing near axis.

no diagonal improvement formula is available for $K=5$, $I=2$.

For $I=1$, $K=5$, the simplest and most satisfactory improvement formula is

$$4\psi_0=\psi_3+3\psi_4 \quad (K=5,\ I=1) \tag{30}$$

(cf. Fig. 1, where point (4) is now on the axis). The formula can be obtained by the method used for (29), but more simply, is an approximation to any function that is even in $\rho$. Again, no satisfactory diagonal improvement formula is available for $I=1$.

*Change of net spacing near axis:* For points $b$ and $c$ of Fig. 6, the previous formulas (9) and (5) will involve negative coefficients. For $b$, formula (30)

$$4\psi_b=\psi_c+3\psi_a \tag{31}$$

is available. For $c$, if a cubic *solution* is passed through $b$, $c$, $f$, $g$, $m$, it is found that

$$8\psi_c=5\psi_b+3\psi_f, \tag{32}$$

which is simple and satisfactory. For $a$, (13) gives

$$39\psi_a=36\psi_b+2\psi_d+\psi_e. \tag{33}$$

It is seen that points $a$, $b$, $c$ may be improved simultaneously as a block, thus speeding up the convergence. New values for $a$, $b$, $c$, are obtained from those at $d$, $e$, $f$ by using successively the following three formulas, which are derived from (31)–(33):

$$\begin{aligned} 7\psi_a&=2\psi_d+\psi_e+4\psi_f,\\ 9\psi_b&=8\psi_a+\psi_f,\\ 8\psi_c&=5\psi_b+3\psi_f. \end{aligned} \tag{34}$$

*Nine-blocks:* The regular nine-block procedure sketched in Section 4 can be used only where the center point has $I\geqslant 4$. Table II gives the coefficients used for improvement of the center point. The formulas for the axis nine-block (Fig. 4) may be obtained by use of (30) for points $b$ and $d$. The center point is given by

$$23\psi_0=15\psi_e+3\psi_f+3\psi_f+\psi_m+\psi_m.$$

The other points are then obtained successively from

$$5\psi_c=3\psi_f+\psi_0+\psi_m,$$
$$4\psi_b=\psi_e+3\psi_0, \quad 4\psi_d=\psi_f+3\psi_c.$$

Blocks centered at $I=2$ or 3 are not available because diagonal formulas are needed for the corner points. Hence it is desirable to arrange blocks along the axis and around $I=4$, 7, etc. (cf. Fig. 7). This leaves the row at $I=2$ to be computed from (29). Since the points at $I=1$ do not enter (29), the values at $I=0$ and 1 are frequently absent from the iteration formulas and may be left for computation after all other values have converged.

## 10. STRESS CONCENTRATION IN A GROOVED SHAFT

In order to illustrate these numerical methods and to show how the various quantities of physical interest may be obtained from the stress-concentration function $\Psi$, a complete solution of the torsion problem for the particular grooved shaft shown in Fig. 7 was carried through.

While we are interested in a shaft of this geometrical shape with arbitrary radius $R$, torque $M$, and shear modulus $G$, it is simplest to carry through the numerical computation for the case of unit radius, unit torque, and unit shear modulus, and then to multiply the results by factors which are determined by the physical dimensions of the quantities concerned, i.e., shear values by $M/R^3$ and angles of twist by $M/R^3G$.

*Stress-Concentration Function.* For the shaft of unit radius and unit torque, $\phi=1$ on the boundary and hence the stress-concentration function $\Psi$ equals $1/\rho^4$ on the boundary and approaches 1 on the straight part of the shaft. It is simpler to compute the function $(\Psi-1)$, which satisfies the same differential equation as $\Psi$ and approaches 0 on the straight parts of the shaft. Figure 8 shows contours of this function. The points used in the finest net and the division into blocks are shown in Fig. 7. The values of Fig. 8 are believed

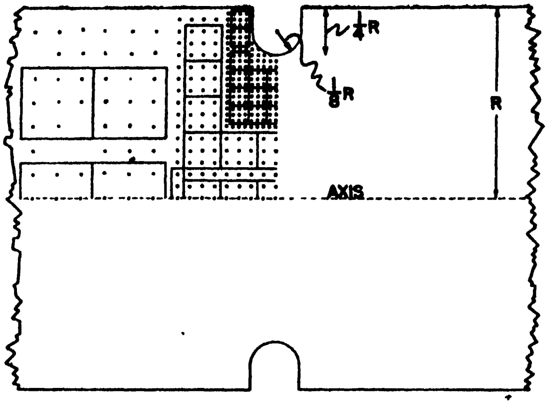


FIG. 7. Grooved shaft showing arrangement of net points and nine-blocks used in numerical stress computation.

to be sufficiently accurate to permit computation of stresses to better than 1 percent accuracy.

*The stress function* $\phi$ for the shaft of unit radius and unit torque is plotted in Fig. 9. Contours are given for $\phi^{\frac{1}{2}}$, which is $\rho\Psi^{\frac{1}{2}}$.

*The twist angle* $\Theta$ is shown in Fig. 10. The values plotted are those of $2\pi\Theta$ for a shaft of unit radius, unit torque, and unit shear modulus. These values were computed from the formula

$$2\pi\Theta(\rho, z) = 2\pi\Theta(\rho, z_0) + \int_{z_0}^{z}\left(4\Psi + \rho\frac{\partial\Psi}{\partial\rho}\right)dz. \quad (35a)$$

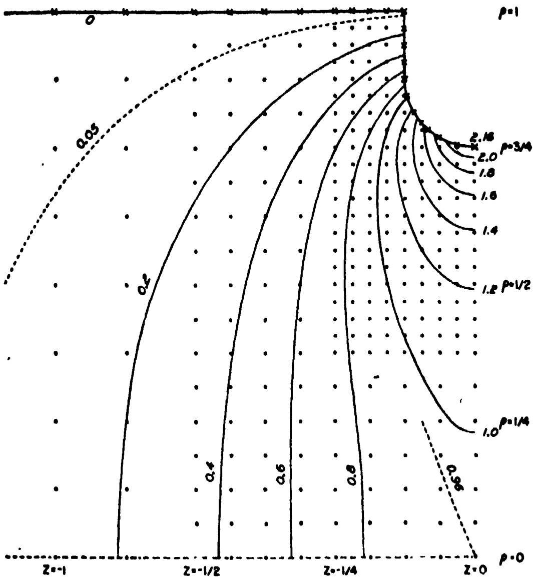


FIG. 8. The stress-concentration function $(\Psi-1)$ for the shaft of unit radius and unit torque.

The value of $\Theta$ was set equal to zero at $z_0 = -1\frac{1}{4}$ (measuring from the plane of symmetry at the groove) where the function $\Psi - 1$ essentially vanishes and the twist is like that of a straight circular shaft. $\partial\Psi/\partial\rho$ was evaluated by the simple parabolic formula

$$\partial\Psi(\rho, z)/\partial\rho = [\Psi(\rho + h, z) - \Psi(\rho - h, z)]/2h$$

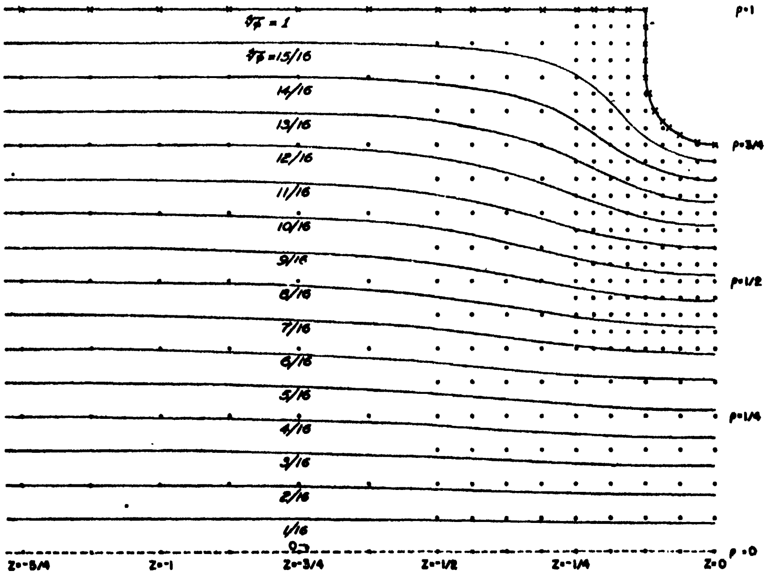


FIG. 9. The stress function $\phi$ for the shaft of unit radius and unit torque. Lines of constant $\phi$ show the directions of the shear vector in the diametral plane.

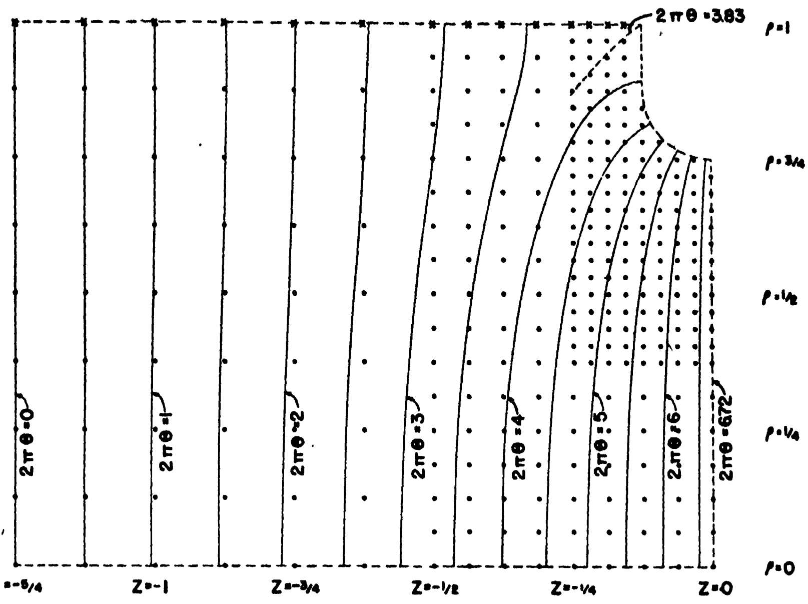


FIG. 10. The twist angle Θ in radians for the shaft of unit radius, unit torque, and unit shear modulus. For the shaft shown in Fig. 7, these values of Θ are to be multiplied by $M/R^3G$. Lines of constant Θ are orthogonal to those of constant $\phi$ shown in Fig. 9, but there is no simple relation between the spacing of the lines on the two plots.

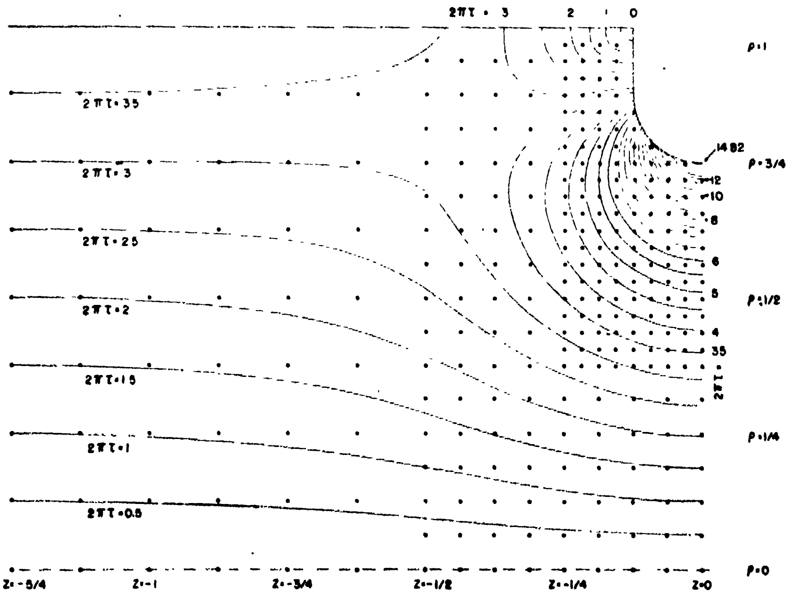


FIG. 11. Magnitude, $\tau = (\tau_{\varphi\rho}^2 + \tau_{\varphi z}^2)^{\frac{1}{2}}$, of the shear vector in the axial plane for the shaft of unit radius with unit torque. The curves of Fig. 9 show the direction of this vector. For the shaft shown in Fig. 7, these values of $\tau$ are to be multiplied by $M/R^3$.

and then the integration was carried out by simple parabolic rules. A check was furnished by the constancy of angle of twist at the plane of symmetry at the groove. Intermediate points were filled in and further checks obtained from the formula

$$2\pi\Theta(\rho, z) = 2\pi\Theta(0, z) - \int_0^\rho \rho\frac{\partial\Psi}{\partial z}d\rho. \quad (35b)$$

It is pointed out that much greater numerical accuracy is obtained in both the angle of twist and in the stresses discussed below by computing them from $\Psi$ rather than from $\phi$.

Figure 10 gives contours[4] of constant twist angle. For a straight shaft without a groove, the contours shown would be uniformly spaced at intervals of $\frac{1}{2}$ in $z$, and the twist at $z=0$ would be $5/2\pi$ radians instead of $6.72/2\pi$ radians for the grooved shaft. The extra twist occasioned by the presence of the groove is thus $1.72/2\pi$ in half the shaft length, or $3.44/2\pi = 0.547$ radian in the whole shaft. For a shaft of radius $R$, torque $M$, and shear modulus $G$, the extra twist introduced by this shape of groove is thus

$$\text{Extra twist} = 0.547M/R^3G \text{ radians},$$

which is equivalent to the twist in a length $0.86R$ of a straight circular shaft of radius $R$. Note that the rate of twist in the section of full diameter decreases as the notch is approached and that this deficiency in twist is made up in the notch in addition to the extra twist given above. One cannot, therefore, determine the extra twist by attaching a torsiometer to the edges of the notch; the value measured will be too great.[4] In our shaft this deficiency amounts to $0.67/2\pi$ radian on the full-diameter part of the shaft near each side of the notch.

Finally, *values of the maximum shear,*

$$\tau = (\tau_{\varphi\rho}^2 + \tau_{\varphi z}^2)^{\frac{1}{2}}$$

are given in Fig. 11. Contours are given for constant values of $2\pi\tau$ for the shaft of unit radius and unit torque. Such values of $\tau$ are to be multiplied by $M/R^3$ in the general case. This plot gives the magnitude of the shear vector in the axial plane, while the curves of Fig. 9 show the direction of this vector.

Values of $2\pi\tau$ were obtained by first computing the components from the formulas

$$\begin{aligned} 2\pi\tau_{\varphi\rho} &= -\rho^2\partial\Psi/\partial z, \\ 2\pi\tau_{\varphi z} &= \rho(4\Psi + \rho\partial\Psi/\partial\rho). \end{aligned} \quad (36)$$

It will be noticed that these formulas involve the same functions of $\Psi$ as in the formulas (35) for twist angle.

The high stress-concentration at the root of the groove is at once apparent from Fig. 11. The maximum stress is

$$2\pi\tau_{\max} = 14.82M/R^3.$$

This is 3.70 times the maximum stress

$$(2\pi\tau_{\max} = 4M/R^3)$$

for a straight circular shaft of radius $R$, and is 1.56 times the maximum stress in a straight circular shaft of radius $\frac{3}{4}R$. This last factor, which compares the maximum stress in the grooved shaft to that in a straight shaft of diameter equal to the root diameter is usually called the *stress-concentration factor* $k$, so for this shaft $k=1.56$. This value is believed to be accurate to within 1 percent.

Our value, 1.56, for the stress-concentration factor may be compared with the experimental value 1.45 read from McGivern's curves of Fig. 4; Bulletin No. 62.[4] The agreement is within the experimental accuracy of McGivern's "Plastic-Flow Method." It may also be compared with Neuber's theory[5] for a "deep external notch." Neuber gives the formula

$$k = \tfrac{3}{4}[1+(s+1)^{\frac{1}{2}}]^2/[1+2(s+1)^{\frac{1}{2}}],$$

where $s$ is the ratio of shaft radius at the root of the groove to the radius of curvature of the

---

[4] This point seems not to have been taken into consideration by J. G. McGivern, Eng. Expt. Station of the State College of Washington, Eng. Bull. No. 58 and 62 (1939), who has made detailed experimental measurements of the torsion of grooved and filleted shafts. We gather from these papers that he placed his twist measuring gauges directly at the edges of the fillets and grooves and thus obtained values of extra twist which were too high. However, the extra twist determined from McGivern's Fig. 5, Bulletin 62, for our particular groove seems to be about twice our computed value, which is a considerably greater discrepancy than can be accounted for by the above comment, but is not inconsistent with the scatter of McGivern's experimental data.

[5] H. Neuber, *Kerbspannungslehre* (Verlagsbuchhandlung, Julius Springer, Berlin, 1937); translated as *Theory of Notch Stresses*, David Taylor Model Basin Trans. 74, Washington (1945).

groove at the root. In our case $s=\frac{3}{4}R/\frac{1}{8}R=6$. This formula is exact for a hyperbolic notch in an otherwise infinite shaft, and is an approximation to any shape of notch deep enough so that the resulting stress-concentration extends significantly throughout the diameter of the shaft at the notch. The value $s=6$ in Neuber's formula gives $k=1.58$, in excellent agreement with our value of 1.56.

We take pleasure in thanking Mr. B. M. Shepard of the Naval Ordnance Laboratory for assistance in the preparation of the figures for the grooved-shaft problem.

## APPENDIX 1. IMPROVEMENT FORMULAS

Consider Fig. 1 and the various cuts in Section 3 showing a point (0) surrounded by four or five neighboring points. The improvement formulas (4)–(11) can be considered as derived in the following way: Let the coordinates of (0) be $z=0$, $\rho=\rho_0$. Write the most general quadric function passing through $\psi_0$ in the form

$$\psi=\psi_0+\alpha_1 z+\alpha_2(\rho-\rho_0)+\beta_1 z^2+\beta_2 z(\rho-\rho_0)+\beta_3(\rho-\rho_0)^2.$$

Now choose the coefficients so that this quadric passes through the neighboring four or five function values $\psi_1$ to $\psi_4$ or $\psi_5$; e.g., for the regular point of Fig. 1, the quadric function passing through $\psi_0$, $\psi_1$, $\psi_2$, $\psi_3$, $\psi_4$ is

$$\begin{aligned}\psi=\psi_0+(\psi_1-\psi_2)z/2h+(\psi_3-\psi_4)(\rho-\rho_0)/2h\\ +(\psi_1+\psi_2-2\psi_0)z^2/2h^2\\ +(\psi_3+\psi_4-2\psi_0)(\rho-\rho_0)^2/2h^2+\beta_2 z(\rho-\rho_0)\end{aligned}$$

with $\beta_2$ arbitrary. Finally, substitute this function in the differential equation (1) and then set $z=0$, $\rho=\rho_0$ to obtain a relation between $\psi_0$, $\psi_1$, $\psi_2$, $\psi_3$, $\psi_4$ which is just the improvement formula (4). This procedure for deriving improvement formulas for various arrangements of points can be formally simplified but is stated in this way to show exactly what is done. *A general quadric is passed through the function values and then required to satisfy the differential equation at just the point to be improved.*

An alternative procedure for deriving improvement formulas is to start with a quantic that satisfies the differential equation (1) *everywhere* and then make it pass through the points of interest in the neighborhood of $\psi_0$. If we are to do this for the points of Fig. 1, a fourth-order polynomial (quartic surface) must be used. The most general quartic *solution* of the differential equation (1) is

$$\begin{aligned}\psi=A+Bz-(1+K)Cz^2+C\rho^2-\tfrac{1}{3}(1+K)Dz^3+Dz\rho^2\\ +\tfrac{1}{3}(1+K)(3+K)Ez^4-(6+2K)Ez^2\rho^2+E\rho^4,\end{aligned}$$

for arbitrary $A$, $B$, $C$, $D$, $E$. Note that this function is automatically even in $\rho$. If now we attempt to put this surface through the four values $\psi_1$, $\psi_2$, $\psi_3$, $\psi_4$ of Fig. 1 in order to determine $\psi_0$, it is found that we must have $E\neq 0$—we cannot put a cubic solution through these values but must use a full quartic. The quartic solution through $\psi_1$, $\psi_2$, $\psi_3$, $\psi_4$ of Fig. 1 determines a unique value of $\psi_0$ and gives the improvement formula

$$\begin{aligned}[96I^3-4K^2I-40KI-48I]\psi_0=[24I^3-6I](\psi_1+\psi_2)\\ +[24I^3+12KI^2-2K^2I-20KI-18I+K^2+7K+6]\psi_3\\ +[24I^3-12KI^2-2K^2I-20KI-18I-K^2-7K-6]\psi_4.\end{aligned}$$

This is a rather complex formula which is probably not so well suited for most numerical work as the simple formula based on a quadric function. Unless the net is fine enough and the function smooth enough so that fourth differences are regular, the use of a quartic function for approximation is apt to introduce rather than to smooth out irregularities. In such numerical work as we are describing it is best to retain simple quadric approximating functions unless one has occasion to reach the point of extreme refinement of net and accuracy.

It is noted that Eq. (28) is an example of the above quartic formula.

## APPENDIX 2. COEFFICIENTS FOR IMPROVEMENT OF NINE-BLOCK CENTER POINT

As a matter of record, we give below the complete expressions, for arbitrary $I$ and $K$, for the nine-block coefficients of which Tables I and II are special cases. If the improvement formula for the center point of Fig. 2 is written in the form

$$\begin{aligned}C_0\psi_0=C_e\psi_e+C_E\psi_E+C_f\Sigma\psi_f+C_F\Sigma\psi_F\\ +C_g\Sigma\psi_g+C_G\Sigma\psi_G+C_m\Sigma\psi_m,\end{aligned}$$

where $\Sigma$ indicates summation of the values at the two points having the same label, the general expressions for the $C$'s are

$$\begin{aligned}C_0=2I\{(512I^4+64I^2(K^2-2K-16)\\ +(K^4-4K^3-60K^2+128K+512)\},\\ C_e=128I^5+128I^4K+4I^3(9K^2-18K-64)\\ +4I^2K(K^2-10K-48)+I(K^4-4K^3-32K^2\\ +72K+128)+K(-K^3+36K+64),\\ C_f=16\{4I^5+4I^4K+I^3(K^2-2K-8)+I^2(-K^2-6K)\\ +I(-K^2+2K+4)+(K^2+2K)\},\\ C_g=32\{2I^5+I^4K-4I^3-2I^2K+2I+K\},\\ C_m=4I\{32I^4+I^2(-K^2+2K-64)+(K^2-2K+32)\},\end{aligned}$$

with $C_E$ obtained from $C_e$, $C_F$ from $C_f$, and $C_G$ from $C_g$ by replacing $I$ by $-I$ and then changing the sign of the whole expression, i.e., $C_E(I,K)=-C_e(-I,K)$, etc.

# Note on a Boundary Value Problem of Reissner and Sagoci

IAN N. SNEDDON
*The Department of Natural Philosophy, the University of Glasgow, Scotland*

(Received June 18, 1946)

## 1. INTRODUCTION

IN their investigation of the torsional oscillations produced in a semi-infinite, homogeneous, isotropic medium by a periodic shear stress applied in an axially symmetrical manner to a circular area of the surface of the medium, Reissner and Sagoci[1] begin by considering a boundary value problem in the theory of elastic equilibrium. The distribution of stress in the interior of a semi-infinite elastic medium is determined when a load is applied to the surface by means of a rigid disk; the torsional displacement is prescribed immediately under the disk, and it is assumed that that part of the boundary which lies beyond the edge of the disk is free from stress. The solution of this mixed boundary value problem is obtained by the introduction of a certain system of oblate spheroidal coordinates.

In the introduction to their paper Reissner and Sagoci point out that the use of Fourier-Bessel methods only reduces the problem to a problem in integral equations which is then reduced to the solution of an infinite number of linear algebraic equations in an infinite number of unknowns. The object of this note is in part to recall that the problem may be reduced to the solution of a pair of dual integral equations.[2] In the static case these dual integral equations reduce to a type which arises in a number of mixed boundary value problems in the mathematical theory of elasticity,[3] and their solution is known. In the present note the dual integral equations are set up by the use of the Hankel transform theorem. It is then shown that the solution for the static case can be derived in this way; though the dynamic problem is not solved a possible approach to it is sketched in Section 3.

## 2. FORMULATION OF THE PROBLEM

We wish to determine the components of stress and of the displacement vector in the interior of the semi-infinite, homogeneous, isotropic, elastic solid $z \geqslant 0$ when a circular area ($r \leqslant r_0$) of the surface is forced to rotate through an angle $\Phi$ about an axis which is normal to the undeformed surface of the medium. It is assumed that the region of the surface lying out with the circle $r \leqslant r_0$ is free from stress. It has been shown by Reissner[2] that, in this case, only the circumferential component $v$ of the displacement vector is different from zero, and that all the components of stress vanish identically except $\tau_{z\theta}$ and $\tau_{r\theta}$ which are given by the relations

$$\tau_{z\theta} = \mu \frac{\partial v}{\partial z}, \quad \tau_{r\theta} = \mu \left( \frac{\partial v}{\partial r} - \frac{v}{r} \right), \tag{1}$$

where, in the absence of internal damping, $v$ satisfies the partial differential equation

$$\frac{\partial^2 v}{\partial r^2} + \frac{1}{r}\frac{\partial v}{\partial r} - \frac{v}{r^2} + \frac{\partial^2 v}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 v}{\partial t^2}, \tag{2}$$

where $c = (\mu/\rho)^{\frac{1}{2}}$. The boundary conditions of the problem are

$$z = 0, \quad v = f(r, t), \quad r \leqslant r_0, \tag{3}$$

$$z = 0, \quad \tau_{z\theta} = 0, \quad r \geqslant r_0. \tag{4}$$

In the case considered by Reissner and Sagoci

$$f(r, t) = \Phi(t) \cdot r. \tag{5}$$

## 3. REDUCTION OF THE PARTIAL DIFFERENTIAL EQUATION BY MEANS OF THE HANKEL TRANSFORM

To solve the partial differential Eq. (2) we introduce the Hankel transform

$$\bar{v} = \int_0^\infty r v J_1(\xi r) dr,$$

of the circumferential component $v$ of the displacement vector. Hence if we multiply both

---

[1] E. Reissner and H. F. Sagoci, J. App. Phys. 15, 652 (1944); H. F. Sagoci, J. App. Phys. 15, 655 (1944).

[2] Such a reduction was first given by E. Reissner in a paper which appeared in Ing. Archiv 8, 229–45 (1937).

[3] See, for example: J. W. Harding and I. N. Sneddon, Proc. Camb. Phil. Soc. 41, 12 (1945); I. N. Sneddon, Proc. Roy. Soc. A, to be published; I. N. Sneddon and H. A. Elliott, Quart. App. Math., to be published.

sides of Eq. (2) by $rJ_1(\xi r)$ and integrate with respect to $r$ from 0 to $\infty$ we have the differential equation

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2}-\frac{\partial^2}{\partial z^2}+\xi^2\right)\bar{v}=0 \tag{6}$$

for the determination of the function $\bar{v}(z, t, \xi)$. Once this solution is known we obtain the expression for the displacement $v$ by means of the Hankel inversion theorem[4]

$$v=\int_0^\infty \xi\bar{v}(z, t, \xi)J_1(\xi r)d\xi \tag{7}$$

from which we have for the stress component $\tau_{z\theta}$

$$\tau_{z\theta}=\mu\int_0^\infty \xi\frac{\partial\bar{v}(z, t, \xi)}{\partial z}J_1(\xi r)d\xi. \tag{8}$$

The arbitrary functions introduced in the solution of the partial differential Eq. (7) are determined by the boundary conditions. The solution must be such that the displacement and both components of stress tend to zero as $z\rightarrow\infty$; when $z=0$ the conditions (3) and (4) must be satisfied. Substituting these relations into the expressions (7) and (8) for $v$ and $\bar{v}$ we obtain the dual integral equations

$$\int_0^\infty \xi\bar{v}(0, t, \xi)J_1(\xi r)d\xi=f(r, t), \quad r\leqq r_0, \tag{9}$$

$$\int_0^\infty \xi\frac{\partial\bar{v}(0, t, \xi)}{\partial z}J_1(\xi r)d\xi=0, \quad r\geq r_0 \tag{10}$$

for the determination of the remaining arbitrary part of the function $\bar{v}(z, t, \xi)$. These equations are equivalent to Eqs. (129) and (130) of Reissner's paper.[2]

### 4. SOLUTION FOR THE STATIC CASE

In the static case the function $\Phi(t)$ of Eq. (5) is a constant, or, in the general case, the function $f(r, t)$ of Eq. (3) is a function of $r$ alone—say $f(r)$. We may therefore take $\partial/\partial t$ to be identically zero in Eq. (7) so that it reduces to an ordinary differential equation with solution

$$\bar{v}=A(\xi)e^{-\xi z}+B(\xi)e^{\xi z}. \tag{11}$$

Since $v$, and hence $\bar{v}$, tends to zero as $z\rightarrow\infty$ we must take $B(\xi)$ to be identically zero in Eq. (11). Substituting this value for $v$ into the Eqs. (10) and (9) we obtain the dual integral equations

$$\int_0^\infty \xi A(\xi)J_1(\xi r)d\xi=f(r), \quad r\leqq r_0 \tag{12}$$

$$\int_0^\infty \xi^2 A(\xi)J_1(\xi r)d\xi=0 \quad r\geq r_0 \tag{13}$$

which are Reissner's Eqs. (129) and (130) with $k=0$. Writing $r=\rho r_0$, $\xi=\eta/r_0$, $\eta^2A(\eta)=F(\eta)a^3$, $f(r)=ag(\rho)$, we see that these equations reduce to

$$\int_0^\infty \eta^{-1}F(\eta)J_1(\eta\rho)d\eta=g(\rho), \quad \rho<1,$$

$$\int_0^\infty F(\eta)J_1(\eta\rho)d\eta=0, \quad \rho>1.$$

Dual integral equations of this type have been considered by Titchmarsh[4] and Busbridge.[5] By making use of the theory of Mellin transforms and known theorems in the calculus of residues Titchmarsh obtained the solution of a generalization of this pair of dual integral equations in the form of a contour integral. The reduction to real form was effected by Busbridge and covers the case arising here; the solution turns out to be

$$F(\eta)=\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\eta\left\{\eta^{\frac{1}{2}}J_{\frac{1}{2}}(\eta)\int_0^1 y^2(1-y^2)^{-\frac{1}{2}}g(y)dy\right.$$
$$\left.+\int_0^1 u^2(1-u^2)^{-\frac{1}{2}}du\int_0^1 g(yu)(\eta y)^{\frac{1}{2}}J_{3/2}(\eta y)dy\right\}.$$

In the case considered by Reissner and Sagoci

$$g(\rho)=\Phi\rho,$$

so that

$$F(\eta)=\frac{4\Phi}{\pi}\left(\frac{\sin\eta}{\eta}-\cos\eta\right)$$

giving finally

$$v=\frac{4\Phi a}{\pi}\int_0^\infty\left(\frac{\sin\eta}{\eta^2}-\frac{\cos\eta}{\eta}\right)e^{-\zeta\eta}J_1(\rho\eta)d\eta,$$

---

[4] E. C. Titchmarsh, *An Introduction to the Theory of Fourier Integral* (Oxford, 1937), p. 240.

[5] I. W. Busbridge, Proc. London Math. Soc. [2], **44**, 115 (1938).

where $\zeta = z/r_0$. Now it is easily shown that[6]

$$I_1 = \int_0^\infty \frac{\sin\eta}{\eta^2} e^{-\zeta\eta} J_1(\rho\eta) d\eta = \frac{\lambda R \sin(\theta+\phi) - \lambda^2 \sin 2\theta}{2\rho} + \tfrac{1}{2}\rho \tan^{-1} \frac{R\sin\phi + \lambda\sin\theta}{R\cos\phi + \lambda\cos\theta},$$

$$I_2 = \int_0^\infty \frac{\cos\eta}{\eta} e^{-\zeta\eta} J_1(\rho\eta) d\eta = \frac{1}{\rho}(R\cos\phi - \zeta),$$

$$I_3 = \int_0^\infty \frac{\sin\eta}{\eta} e^{-\zeta\eta} J_1(\rho\eta) d\eta = \frac{1}{\rho}(1 - R\sin\phi),$$

$$I_4 = \int_0^\infty \cos\eta\, e^{-\zeta\eta} J_1(\rho\eta) d\eta = \frac{1}{\partial}\left\{1 - \frac{\lambda}{R}\cos(\theta-\phi)\right\},$$

where the quantities λ, $R$, $\theta$, and $\phi$ are defined by the relations

$$\lambda^2 = 1 + \zeta^2, \qquad \zeta \tan\theta = 1,$$
$$R^4 = (\rho^2 + \zeta^2 - 1)^2 + 4\zeta^2, \quad 2\zeta \cot 2\phi = \rho^2 + \zeta^2 - 1.$$

Hence we have the formulae

$$v = \frac{2\Phi a}{\pi\rho}\left[\lambda R \sin(\theta+\phi) - \lambda^2 \sin 2\theta + \rho^2 \tan^{-1}\frac{R\sin\phi + \lambda\sin\theta}{R\cos\phi + \lambda\cos\theta} - 2R\cos\phi + 2\zeta\right],$$

$$\tau_{z\theta} = \frac{4\Phi\mu R}{\pi\rho}\left[\sin\phi - \frac{\lambda}{R}\cos(\theta-\phi)\right],$$

for the determination of $v$ to $\tau_{z\theta}$ at any point in the interior of the elastic medium; a similar expression may be derived for the other shear stress $\tau_{r\theta}$. When $z=0$ and $\rho>1$,

$$I_1 = \frac{1}{2\rho}\left[(\rho^2-1)^{\frac{1}{2}} + \rho^2\left\{\frac{\pi}{2} - \tan^{-1}(\rho^2-1)^{\frac{1}{2}}\right\}\right],$$

$$I_2 = (\rho^2-1)^{\frac{1}{2}}/\rho,$$

and when $z=0$, $\rho<1$,

$$I_3 = \frac{1}{\rho}[1 - (1-\rho^2)^{\frac{1}{2}}], \quad I_4 \doteq \frac{1}{\rho}[1 - (1-\rho^2)^{-\frac{1}{2}}],$$

giving for $v$ and $\tau_{z\theta}$ when $z=0$,

$$v = \Phi a \rho\left[\left\{1 - \frac{2}{\pi}\tan^{-1}(\rho^2-1)^{\frac{1}{2}}\right\} - \frac{2}{\pi}(1 - 1/\rho^2)^{\frac{1}{2}}\right], \qquad \rho>1,$$

$$\tau_{z\theta} = -\frac{4\Phi\mu}{\pi}(1/\rho^2 - 1)^{-\frac{1}{2}}, \quad \rho<1$$

in agreement with the expressions found by Reissner and Sagoci.[7]

---

[6] See, for instance, I. N. Sneddon, Proc. Camb. Phil. Soc. 42, 27 (1946).

[7] Reference 6, Eqs. (23) and (24).

---

# Erratum: Theory of Filler Reinforcement

[J. App. Phys. 16, 55 (1945)]

JANE M. DEWEY
*United States Rubber Company, Passaic, New Jersey*

THE last line of Eq. (3) should read

$$-\frac{1}{\gamma-1}\left[\frac{\partial P_{\gamma-2}^{|m|}}{\partial\theta} i_\theta + \frac{im}{\sin\theta} P_{\gamma-2}^{|m|} i_\phi\right].$$

In Eq. (5)

$$M = -\tfrac{1}{3}(A+B-2C)P_2(\cos\theta) + \tfrac{1}{12}(A-B)P_2^2(\cos\theta)(e^{2i\phi} + e^{-2i\phi}).$$

In Eq. (6)

$$b_{-2} = (3\lambda_1 + 2\lambda_2 - 3\lambda_1' - 2\lambda_2')(4\lambda_2 + 3\lambda_1' + 2\lambda_2')^{-1}R^3,$$
$$c_{-2} = -5(3\lambda_1 + 5\lambda_2)DR^3,$$
$$c_{-4} = 9(\lambda_1 + \lambda_2)DR^5,$$
$$a_1 = (A+B+C)(\lambda_1 + 2\lambda_2)(4\lambda_2 + 3\lambda_1' + 2\lambda_2')^{-1}.$$

Journal
of
Applied Physics

Volume 18, Number 2 February, 1947

# Physics in 1946

BY PHILIP MORRISON
*Cornell University, Ithaca, New York*

NINETEEN forty-six was the first year of peace. Throughout the world, and especially in the United States, the generally difficult problems of reconversion began in earnest. Reconversion meant much for physics and physicists. The strange symbols of wartime work—the SCR-584 radars, the HVAR rockets, the VT fuzes, the compounds of tuballoy—began to lose their compelling interest and their jealously-guarded secrecy. The great "amateur" wartime laboratories, where academic physicists and their industrial and engineering colleagues had worked so hard and so successfully on war program, began to dwindle. The publication of volumes of reports began. And physicists went back, a little rustily, to the problems of the days before the war. Most of those problems were still there, for not much fundamental progress had been made during the war years. But gone was the reluctance to do big things, gone the sometimes valuable, sometimes hampering isolation of the research worker. Physics, especially nuclear physics and its related frontier fields, had grown up. It was pretty well organized, in the wake of the OSRD and the Manhattan Project.

The work of the year very much reflected the problems of the physicist. Much work was begun, but not very much completed. People had plans, often great and exciting ones, but still only plans. Some wits began to talk of the latest "Physical PREview." Teaching loads were heavy, and the flood of students, good and serious and deserving students, drew many research workers away from problems they had almost begun to remember. New laboratories had sprung up, like Oak Ridge and the Argonne, and in the first postwar year began to make their mark on public research.

Most striking of all was the essential disappearance of the peacetime international community of physics. The United States, tired enough by years of war but still incredibly rich by the standards of Europe and Asia, was pre-eminently the home of physics. Our once good communication with the rest of world had atrophied with the secrecy and the contingencies of wartime. By the end of the year, only a beginning had been made towards a return of the days when physics was truly international. Few physicists hoped more wistfully for anything than for the full restoration of the ways of peace in travel, in publications, and in the spirit of a world-wide science.

The present account is for all these reasons a fragmentary account, mainly of "work in progress," and mainly of work here in America. But it is work of the highest promise for physics, work carried on with the highest hope that in 1947 and the years ahead the best of the old spirit will come to employ the great new tools which are the legacy of war.

## SUPERCONDUCTIVITY

One of the most spectacular results of the year was in the tradition of the world before the war. Professor Richard A. Ogg, working essentially alone in the Chemistry Department of Stanford University, discovered a substance which is electrically superconducting at liquid-air temperatures and even above.[1] This statement has little in common with the dramatic discoveries of new particles, but this work may be as fundamental a step towards the understanding of the strange phenomenon of superconductivity as was the finding of the neutron to nuclear physics.

For decades it has been known that some metals and alloys cooled to very low temperatures lost all electrical resistivity. Resistances of many ohms dropped by twelve or fourteen orders of magnitude when the transition temperature was reached. Strange electrical and thermal experiments could be performed with such extraordinary solids. By moving a ring of the material out of a magnetic field in which it was cooled, a current was induced. This current died away with a time constant given by the inductance and the extremely small resistance of the cold metal. It is told that the Leyden laboratory workers delighted to prepare such persistent currents, and to send the loop, still cold in its Dewar, to a distant place where the lecturer could demonstrate that the current was still flowing, hours or even days after it had been started up! Such conductors demonstrated Lenz's law with a vengeance: their low resistivities meant that the induced currents cancelled the external fields to a T, and magnetic fields thus cannot be set up within superconductors. They are almost perfectly diamagnetic.

The large scale theory of such phenomena is fairly clear, but it is quite unsure what is the electronic and atomic mechanism by which the state of superconductivity is caused. Only one thing seemed essential: extremely low temperatures. No superconductor was known which showed the effect above ten degrees absolute.

Ogg had long been interested in a famous and strange set of substances, the solutions of the alkali metals in liquid ammonia. For twenty years these materials have provided work for physical chemists concerned with the nature of ionization in solution. The example which proved so interesting was the dilute solution of sodium in ammonia, in concentrations of twenty or thirty grams of metal per liter of solvent. That solutions of this kind are sufficiently strange in their other properties was well known. They look like metals, with a quickly-frozen solution having a metallic steely blue luster. At −50°C or so, the homogeneous solution separates in two differently-colored phases, bronze and blue, if slowly cooled. It is not hard to prepare the material. In one method, an annular glass trap is employed, attached to a vacuum line.[2] Water and air must of course be kept out. The sodium metal is placed in the system, and anhydrous ammonia brought into contact with the metal. The system is kept cool, below the boiling point of the ammonia, at −34°C, and the solution forms. To demonstrate the superconductivity it is enough to place the liquid solution in its vessel inside a solenoid coil, and to freeze it in liquid air within a matter of ten seconds or so. When the frozen ring is removed from the liquid air, it is placed near a flip coil, and the magnetic field of the ring detected by a galvanometer deflection. A magnetic field will repel the ring strongly, because of the very large *negative* susceptibility of the substance. Such tests seem to prove the superconducting nature of the substance.

It is not quite so simple as that, as usual. Out of a few score preparations, only a few are successful. Apparently the large volume change on freezing causes the solid to crack seriously, and the resistance is the resistance of the cracks and not of the solid itself. One must be patient to find the effect. Two attempts[3,4] to confirm Ogg's work failed to do so, probably for this reason, but it has been confirmed by at least one other worker.

Why the interest? One more superconductor for the handbooks does not sound exciting. But there is a vast difference. This material is a superconductor not at 5 or 10 degrees absolute temperature, but at more than 90°K. Recall that thermodynamically the gap between these tem-

---

[1] R. A. Ogg, Phys. Rev. 69, 243 (1946).

[2] J. W. Hodgins, Phys. Rev. 70, 568 (1946).

[3] Boorse, Cook, Pontius, and Zemansky, Phys. Rev. 70, 92 (1946).

[4] Daunt, Désirant, Mendelssohn, and Birch, Phys. Rev. 70, 219 (1946).

peratures is much greater than the gap between the temperature of liquid air and room temperature. Consider, for instance, the amount of work required to remove a given amount of heat energy, say by the use of a mechanical refrigerator. If the region to be cooled is at liquid-air temperature and the condenser of the refrigerator at room temperature, less than one-third as much work is required as if the heat had to be moved from a region at 10°K to a condenser cooled by liquid air. Thus Professor Ogg's solutions hold the suggestion that a superconductor can perhaps be made at room temperature. One cannot predict such a result, of course, but it would be a foolhardy statement now to deny its possibility. A little speculation on what would happen to laboratory apparatus and indeed to all electrical devices if fairly large current density could be achieved *without appreciable ohmic loss* is worth while. Even if no such near-fantastic result ever ensues, the first superconductor at temperatures so "high" as those of liquid air is sure to prove a challenge and a stimulus to the theory of this strange phenomenon.

Ogg predicted this property of the sodium-ammonia system from his own theory[5] of the structure of this unusual substance. His theory is formally strange and unconventional to most physicists, and is unconvincing to this author at least. It involves the notion of paired electrons which move in cavities in the solvent, after escaping from the valence bonds of the alkali metal. These paired electrons have no resultant spin and satisfy a condition years ago proposed to explain superconductivity as an effect bound to occur with particles which have no spin. The stability of these arrangements and the order-of-magnitude of concentrations and temperatures do not appear evident from Ogg's so far preliminary theory. It is to be hoped that the next year will bring wider interest in this remarkable problem both from the experimental and the theoretical side. There is hardly a laboratory which cannot make a superconductor now! Liquid hydrogen generators and the rest of the special techniques of cryogeny are no longer needed.

### THE PUSH TOWARD HIGH ENERGY

The one most characteristic feature of nuclear physics is the high energy per atom involved in the reactions. Where chemical reactions release a few electron volts for each atom involved, the reactions of nuclear physics release or require millions. The very fact of such high specific energies makes it possible to do experiments in which only individual atoms take part, observability following only from the highly energetic nature of the motions and transformations of the reactants. Nuclear physics—with of course the spectacular and decisive exception of the chain reaction of uranium—does not often work with macroscopic amounts of matter.

But to give even single atoms the millions of electron volts they need often requires large scale apparatus. The engineering which is so typical of such laboratories today began in the first efforts to concentrate a few million volts in ions of hydrogen and heavy hydrogen. Since Rutherford's Cavendish days, this has been done by simply setting up a few million volts potential difference in the laboratory (with a.c., as in a surge transformer set, with electrostatically-generated d.c., as in the familiar Van de Graaff machines) and letting the ions fall through the potential difference in a vacuum tube. Always the demand was for higher energy.

The first energy barrier which it was sought to overcome was the electrostatic barrier, the energy required to bring one positively-charged projectile very near the positively-charged nucleus. One wanted the charges near enough so that the nuclear matter could touch and "stick," to produce new radioactive nuclei and to initiate nuclear reactions. The size of this barrier is easy to compute: the energy of repulsion is $Zze^2/r$, where $ze$ is the charge of the projectile (generally $e$ or $2e$ for protons, deuterons, or alphas), $Ze$ the charge of the target nucleus, ranging to $94e$ for the heaviest known, and the distance $r$ is the nuclear radius, never much larger than $10^{-12}$ cm. The answer turns out to be some ten million electron volts needed to cross the barrier for the most difficult cases. No one has yet established a potential difference above six million or so. But the desire to produce particles with such energies has led to ingenious solutions.

---

[5] R. A. Ogg, Phys. Rev. 69, 668 (1946).

The most famous of these is the cyclotron, in which a modest radiofrequency potential difference of a hundred or two kilovolts is applied again and again to the moving particle which is bent into a spiral path by a large magnetic field. Successive passages across the r-f drop occur uniformly spaced in time and, "resonant" with the r-f frequency, lead to high final energies. The Crocker cyclotron at Berkeley, the largest machine of its kind, is capable of making forty million-volt alpha-particles, which cross any nuclear barrier with energy to spare, and make possible many complicated reactions. Why go further?

The answer is, of course, that there are still more remote objectives. The probing of nuclear matter by bringing the "test charge" which is the rapidly moving projectile closer and closer to the nucleus has given us information about distances of the order of $10^{-12}$ cm or somewhat less; we would like to look at matter even more intimately. The closer the impact, the more energy transferred, the finer the detail of exploration. We learned many years ago that there were particles of very short life, which we cannot find in our laboratories, but which stream in to the earth in the cosmic rays. These objects, known now as mesotrons, because of their mass of 200 electron masses (intermediate between that of electron and proton), are believed on quite general grounds to be associated with those extraordinary attractive forces which make nuclear matter sticky, which bind together the neutrons and protons of the nucleus. The argument for the connection is essentially this: The electromagnetic forces are by now familiar. They are long range forces, falling off as the inverse square of the distance between two interacting particles. Now, these forces, which propagate, of course, with the finite velocity of light can be thought of as caused by the transfer of quanta between the moving charges. This is quite evident for the transverse electromagnetic waves, but it is true at least formally for the static Coulomb field as well. From this point of view electric charges are surrounded by quanta. If there is energy available, these quanta may be set free to go as far as they will. Otherwise they must be emitted, last a very short while, and then be reabsorbed by the particle which emitted them or by another particle nearby. By analogy the same mechanism is used to account for nuclear forces by the exchange of mesons, It turns out that the Compton wave-length, $\lambda = h/mc$, of the particles transferred determines the effective range of the forces. Thus, for quanta, the rest mass is zero, and the "Compton wavelength" infinite, and the range infinite. The force decreases only for geometrical reasons, and is simply inverse-square. But the nuclear forces act only at short range, about $2 \times 10^{-13}$ cm. The rest mass corresponding to such a range is about two hundred electron masses, just that observed for the mesons of cosmic rays. The inference is strong that around nuclear particles there is a transient cloud of mesons, and that if one supplies enough energy by collision or even by energetic "light" quanta, such mesons can be set free. The study of this mechanism would correspond for the nuclear forces to the study of Maxwell's equation, and could lead to the understanding of the nuclear forces in detail. It must be said that since the first suggestion of Yukawa in 1936 that such mesons might exist, reinforced by their discovery in the cosmic rays two years later, the best theorists have worn thin their patience on this theory. No consistent description has yet been given of the properties of nuclear forces. All the more has this challenged experimenters to make mesons in the laboratory and there to study them in the number and the detail which is impossible while their source is still only the cosmic ray.

How much energy does one need? No one knows. There are clues. Certainly at least 100 Mev is needed, just for the rest energy, known to be 200 times the 0.51 Mev which is the electron rest energy. Perhaps they can be made only in positive and negative pairs, as are electrons in the field of a nucleus. Then at least 200 Mev is needed. More will be required to make them in some quantity, for one must do more than merely tickle the threshold of the reaction. So a popular target for the ingenious builders of machines has been 300 Mev. There are some who think that many mesons must be made at once really to study the details of their creation; these pessimists (or optimists) are planning in billions of volts, but still for some years in the future!

The first postwar year saw active construction on large machines and smaller prototypes, all shooting at the 300-Mev target, or thereabouts.[6] Most direct attack was that planned long ago at Berkeley, where a giant cyclotron had been built. This seventeen-foot magnet was built in 1940, and used during the war in the experiments which led to the construction of the great electromagnetic separation plant at Oak Ridge. By simply extending the cyclotron design to real California scale, 200-Mev deuterons were expected, somewhat short of the popular goal, the more because such particles directed at nuclear particles have an effective velocity not much more than half of that in the laboratory system of reference. The transferrable energy is given by the velocity with which the particles are brought together computed in the frame of reference in which their center of gravity is at rest. Even this energy is most difficult to get with a cyclotron. In the cyclotron, the particles must all spiral out in step, at constant phase, returning in each circuit to the region between the *D*'s, where the high r-f voltage is applied. The familiar resonance condition is that $2\pi f = eH/mc = \omega$, where $e$ and $m$ are the charge and relativistic mass of the particle, i.e., $m = \frac{(\text{rest energy} + \text{kinetic energy})}{c^2}$, $H$ the constant magnetic field, $f$ the frequency of the r-f oscillator and $\omega$ the angular velocity of the particle. This implies that the particles at all radii have the same resonant frequency, though some go fast and others slow. This is the key to the success of the machine. But the mass value, hence the resonant frequency, is not, in fact, constant, if the relativistic variation of mass with velocity be considered For 200-Mev deuterons this is not a small effect. The hope was, however, that by running up the highest possible r-f voltage, several million volts, some particles could be gotten out in spite of their having fallen a little out of phase on each step, for the number of times they need circle in the machine is only a hundred. The giant cyclotron has indeed been operating at Berkeley since November 1, producing its deuterons as planned, but the principle of operation is quite changed, and the deuterons make not a hundred circles in their spiral path, but nearer to ten thousand. And the r-f oscillator is not a giant device producing thousands of kilowatts but an oscillator which would be considered small even for an ordinary-sized cyclotron.

The story of this change is also a story of 1946. It began, of course, much earlier. The start was probably the development of the betatron in which the accelerating voltage is not applied by any electrodes, but is simply the induced e.m.f. caused by the changing magnetic flux in the a.c. magnet, acting on the electrons in their vacuum doughnut as if they were the secondary turns of a transformer. The total induced voltage in present design is only some seventy-five volts per turn, but the electrons come out with a hundred million volts in the General Electric betatron. These particles have circled the primary about a million times, traveling something like a thousand miles in the vacuum tube. The key to this machine is evidently the stability of the electron orbits, so that a chance disturbance from stray fields or by collision with a gas molecule does not throw too many electrons against the glass walls of the vacuum tube. This is a good machine for energies up to several hundred million volts. The 100-Mev General Electric model produced the first man-made radiation of such high energy, but, in spite of hopes and even the illusion of success, probably has made no mesons. Professor Kerst is now engaged in constructing a 300-Mev model, of quite advanced design, at the University of Illinois in Urbana.

In 1945 Professor E. M. McMillan, then at Los Alamos, of the University of California, proposed a new type of acceleration. Actually, the same proposal had been published months earlier and completely independently by the Soviet physicist V. Veksler. The McMillan-Veksler idea[7] stems in a way from the demonstration in the betatron that it was not foolhardy to plan for very long paths in the vacuum tube, if the conditions of stability are properly fulfilled by the design of the machine. The fact that one could count on geometrical stability had been proved in the betatron. McMillan began to think of the use of radiofrequency accelerating electrodes, as in the cyclotron, with stability in

---

[6] L. Schiff, Rev. Sci. Inst. 17, 6 (1946).

[7] E. McMillan, Phys. Rev. 68, 143 (1945); V. Veksler, J. Phys. (USSR) 9, 153 (1945).

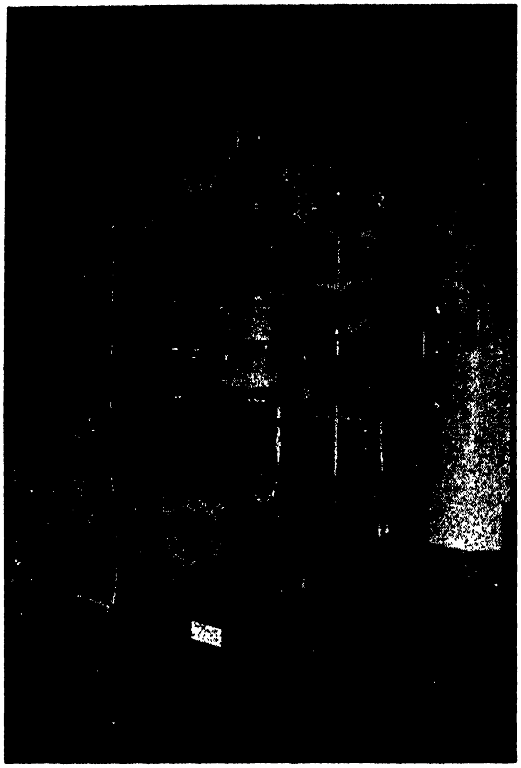


FIG. 1. The first synchrotron to operate in America. This device is installed at the General Electric Research Laboratory at Schenectady, and is under the supervision of Dr. H. C. Pollock. The vacuum tube is visible between the magnet poles.

time, or, as he calls it, phase stability. With such a condition, the resonance condition of the cyclotron might be far from satisfied by all particles, and yet eventually after many turns, the particles might gain a large energy.

For definiteness, think of a particle in the cyclotron at perfect resonance. Each time it reaches the gap between the accelerating electrodes the r-f voltage has reached its maximum. If the particle happened to arrive a little early, the r-f voltage has not yet climbed to its maximum and the particle gets a little less energy than it should have gotten. It will then make the next turn in the field with a somewhat too high angular velocity (the equivalent mass is too small) and arrive still earlier next time. Such a particle lacks phase stability and will gradually fall more and more out of step. It will never gain much energy. Now consider a very different case, that of a particle which arrives when the r-f voltage drop across the gap between electrodes—the accelerating field—has fallen to zero. This particle is just 180 electrical degrees out of phase with the r-f voltage. It will be in resonance, but it will gain no energy in passing the gap. There are two such nodes in the complete cycle. Let us consider the one in which the voltage across the gap is changing from being slightly accelerating through zero to be slightly decelerating. (The other node will not provide stability.) Now suppose the out-of-phase particle arrives just a bit early. It will feel a small amount of an accelerating electric field. It will gain energy, its equivalent mass will increase, its angular velocity go down, it will begin to lag, and next time it will be more nearly 180 degrees out of phase again. If it came a bit late, it would cross a small decelerating potential drop, and be reduced in energy. The mass would go down, the angular velocity go up, and it would come more closely to the 180 degree phase. A particle exactly out-of-

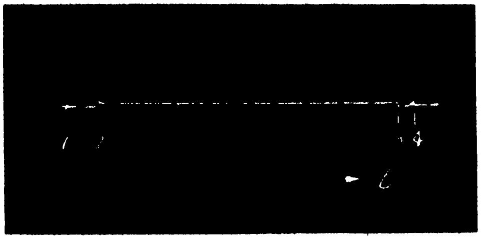


FIG. 2. An oscillogram showing operation of the 75-Mev GE synchrotron. The trace is cut off every 100 microsec. (1) marks the time of electron injection. At this time the machine is operating as a betatron. At (2) the r-f voltage is turned on. The electrons have by now about two-Mev energy. At (3) the r-f is turned off and at (4) is a signal from a photo-multiplier tube placed in the x-ray beam. For this trace the energy was about 20 Mev.

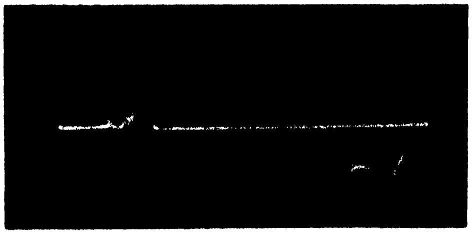


FIG. 3. Another oscillogram of synchrotron operation. The length of the sweep is 1000 microseconds. The trace begins at the time the r-f was turned off, and the electrons soon begin to strike the internal target. The irregular signal shown is taken from the output of a photo-multiplier tube placed in the emergent x-ray beam. Note the complex structure of the beam. This effect is still unclear. The machine is operating between 60 and 70 Mev.

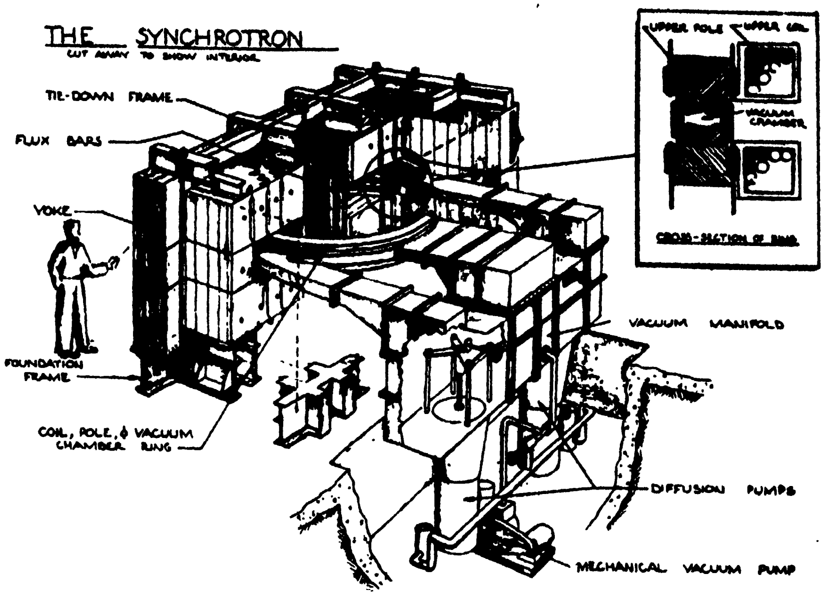


FIG. 4. The synchrotron being built at Berkeley for 300-Mev electrons by the group under Professor E. M. McMillan. The flux bars referred to bars of iron which saturate at a modest field. Until they saturate they strengthen the central field of the machine so that the electrons will accelerate under betatron conditions. When they saturate the electrons are moving at a modest energy, and the synchrotron r-f voltage is turned on.

phase then will have phase stability; it will tend to maintain this condition both at low and at relativistic velocities. Now suppose either the magnetic field or the alternating frequency be slowly increased, changing a little, but only a little, during the time the particle is making its turn in the magnetic field. The stability in phase will act to insure that the particle increases its mass to keep resonant with the increased frequency. The particle will cross the gap just enough before the time when the electric field is zero so that it gains energy sufficient to keep it near resonance. As long as the frequency or magnetic field changes slowly, the particle will oscillate in phase back and forth across the 180 degree time, gaining on the average just enough energy to be resonant under the new conditions. Thus after many turns, picking up a very small amount of the available accelerating energy on each turn, the particle can reach high energy. The frequency may be varied with constant magnetic field, and the particle will spiral out as its energy increases. Or the r-f frequency and orbit radius may stay constant or nearly so and the magnetic field increased. Or combined changes may be made. In all such devices the particles will not be accelerated in a steady stream as in a cyclotron, but in pulses, repeating as rapidly as the frequency is varied or the magnetic field made to increase.

These devices, of which there is clearly a large family, are generically called "synchrotrons." The name is based on the analogy between the motion of the particles and that of the rotor of a synchronous motor. The rotor spins at exactly the synchronous speed with no load. But loading the shaft does not change the speed. The phase slips behind far enough so that the field differences will supply the needed energy to the load. In the same way the particles in the synchrotron slip out of exact phasing, "hunting" in fact for the node, but gaining energy just enough to compensate for the changing frequency or magnetic field.[8]

The first American synchrotron completed is a seventy-five million volt electron machine, Fig. 1. The pulses of fast electrons from this machine are shown in Fig. 2. A synchrotron was made even earlier in Britain[9] by placing a small electrode for the r-f voltage inside a betatron doughnut. The output of the machine went from four millions to more than double that, since the saturation of the central part of the magnet core in the betatron limits the field at the orbit to less than half of practical design saturation values. The stability of orbits in the betatron demands that the central flux change be larger by a

---

[8] D. Bohm and L. Foldy, Phys. Rev. 70, 249 (1946); H. C. Pollock, Phys. Rev. 69, 125 (1946).

[9] Goward and Barnes, Nature 158, 413 (1946).

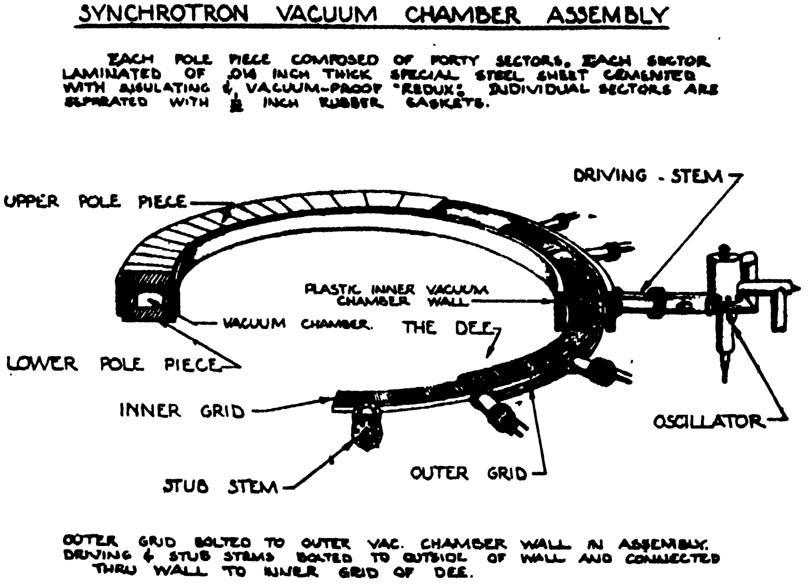

SYNCHROTRON VACUUM CHAMBER ASSEMBLY
EACH POLE PIECE COMPOSED OF FORTY SECTORS. EACH SECTOR LAMINATED OF .014 INCH THICK SPECIAL STEEL SHEET CEMENTED WITH INSULATING & VACUUM-PROOF "REDUX". INDIVIDUAL SECTORS ARE SEPARATED WITH $\frac{1}{8}$ INCH RUBBER GASKETS.
DRIVING - STEM
UPPER POLE PIECE
PLASTIC INNER VACUUM CHAMBER WALL
VACUUM CHAMBER.
THE DEE
LOWER POLE PIECE
INNER GRID
OSCILLATOR
OUTER GRID
STUB STEM
OUTER GRID BOLTED TO OUTER VAC. CHAMBER WALL IN ASSEMBLY. DRIVING & STUB STEMS BOLTED TO OUTSIDE OF WALL AND CONNECTED THRU WALL TO INNER GRID OF DEE.

FIG. 5. The vacuum assembly for the Berkeley synchrotron. Note the r-f electrode "dee." It is constructed as a grid instead of as a solid electrode to reduce eddy-current losses and fields which would be set up by the strong a.c. magnetic field. The r-f energy is supplied by a pulsed water-cooled oscillator at about 48 megacycles. The stubs are provided to adjust the position of the nodes on this rather high impedance structure.

definite factor than that at the orbit. Synchrotron stability does not require any such condition, for the voltage gain comes from the r-f field imposed by electrodes, and not from the transformer effect except incidentally. The GE machine actually begins the acceleration of its electrons as a betatron; then, when they have reached an appropriate starting energy, the r-f is turned on.

The giant cyclotron is working now accelerating heavy particles not as a conventional cyclotron, but as a synchrotron, or as it is sometimes called, a synchro-cyclotron. Frequency modulation of the r-f supply accomplished by a rapidly rotating condenser in the plate tank of the r-f oscillator brings out the high energy beam, after thousands of turns, with a modest amount of r-f power. The synchrotron principle maintains resonance acceleration in spite of the relativistic variation of mass.[10] The synchro-cyclotron accelerates deuterons, which are caused so far to strike an internal beryllium target, producing a beam of very fast neutrons. The first operation of the machine was on schedule, November 1, 1946, and experiments with the new fast neutrons are already in progress. No mesons have

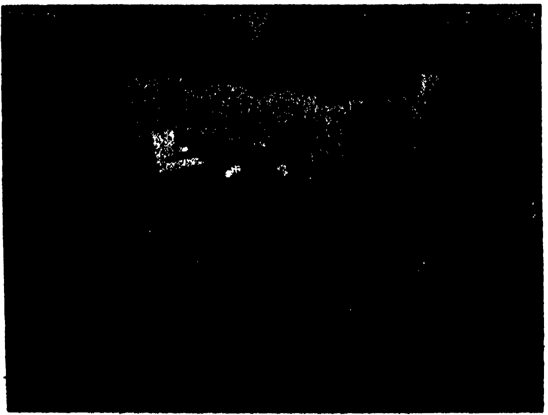


FIG. 6. Photograph of a model of the Berkeley synchrotron showing the driving stem and stubs.

FIG. 7. The other side of the synchrotron model, showing the vacuum pumps in their pit, and the large vacuum manifold.

---

[10] Richardson, MacKenzie, Lofgren, and Wright, Phys. Rev. 69, 669 (1946).

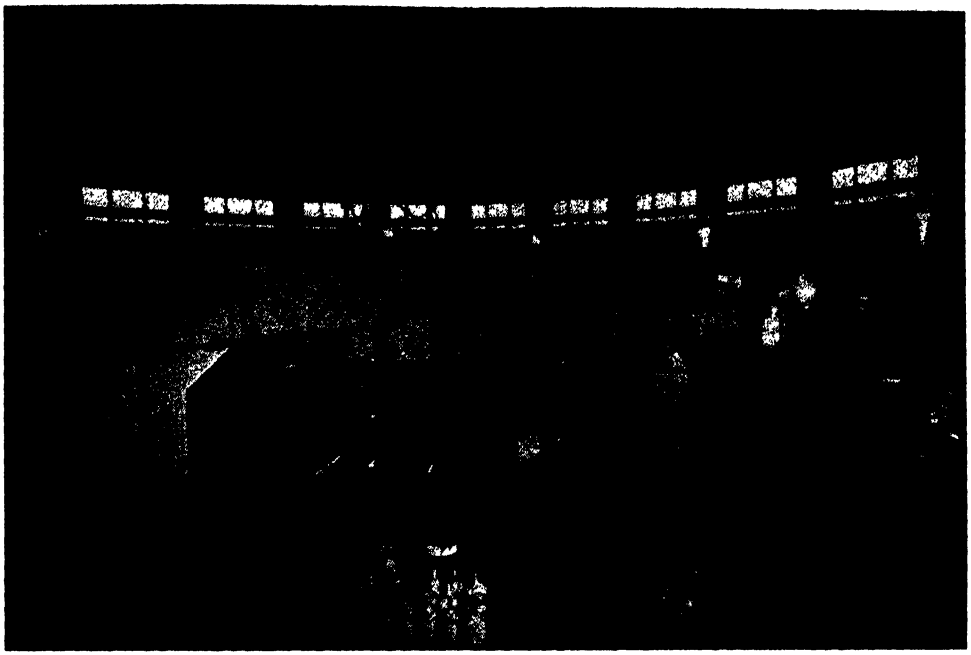


FIG. 8. A general view of the giant Berkeley cyclotron, now operating for 200-Mev deuterons as a synchro-cyclotron. Note the large radial crane of thirty-ton capacity overhead, the magnet with its 184" pole pieces (there is a man visible against the right leg of the magnet yoke), the two 32" oil diffusion pumps used for evacuating the vacuum "can," whose thick iron walls are visible between the pole pieces, and the rails, to the left of the pumps, used for removing the ion source. The temporary concrete block shielding wall and a cloud-chamber set-up are seen at the right. (The photographs of this machine were furnished through the kindness of Professor R. L. Thornton, in charge of the giant cyclotron group at Berkeley.)

yet been seen; but important results on the nature of nuclear forces seem sure already (December 15, 1946). It is planned to bring the deuteron beam itself out of the machine in the near future.

Michigan,[11] Massachusetts Institute of Technology, Cornell, the General Electric Laboratory, Berkeley, and no doubt other laboratories are now engaged in design and construction of synchrotron-type electron accelerators of the 300-Mev size, in a variety of ingenious exemplifications of the same principle. The mechanical and electrical engineering problems involved in such work are not small, and many a design variant may have its own special advantages. General problems like starting the particles in their orbits, removing the particles from the machine after acceleration, detection methods specially suited to these pulsed sources, and many others may be expected to build up a whole science of very high energy physics in the next year or two.

One simple and elegant achievement of the year in this field was the first successful removal of a high energy electron beam from the Urbana betatron, at about 20 Mev.[12] The device which did the work is simply a piece of iron shaped like a U-channel, mounted in the vacuum doughnut in the right position. The electrons enter the U-shaped slot, where the iron walls shield them from most of the magnetic field, and shoot out in the field-free space in a straight line to a target or a thin window outside the machine. The success of this device should make possible many experiments with fast electrons, up till now never available in a well-directed beam.

The idea of bending the fast particles in circular orbits and causing them to retrace their steps is evidently economical. But if the magnet grows

---

[11] H. R. Crane, Phys. Rev. 69, 542 (1946); Phys. Rev. 70, 800 A (1946); Pollock *et al.* Phys. Rev. 70, 798 (A) (1946).

[12] Skaggs, Almy, Kerst, and Lanzl, Phys. Rev. 70, 95 (1946).

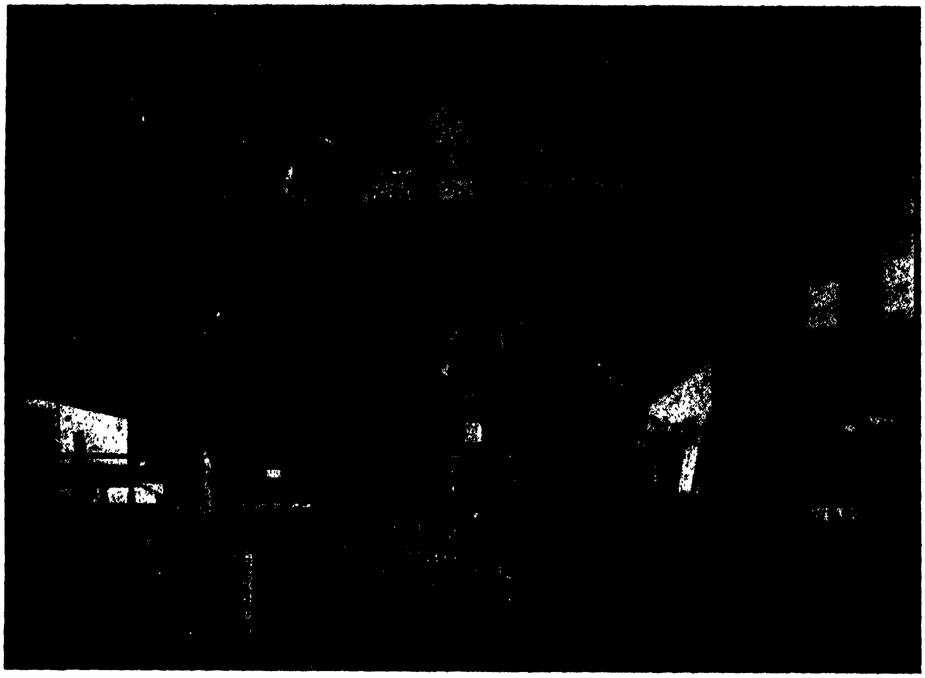


FIG. 9. A view of the big cyclotron from the side opposite to that shown in Fig. 8. In the right foreground, on the truck which moves on rails in the floor, are mounted the round vacuum housing for the rotating condenser, its associated vacuum pump, behind them the oscillator housing. The vacuum chamber is here shown clearly in operating position in the field. To the left may be seen the horizontal rod, which extends into the vacuum chamber holding the internal beryllium probe against which the beam is made to strike. Neutrons were produced at 100 Mev or more on November 1, 1946.

to be very large the economy of the scheme begins to seem less obvious. Perhaps it is better to let the particles fly in a straight line down a long tube. If one can set up a certain electric field strength, the cost of the linear accelerator goes up about as its length (the parts are simply repeatedly placed end-to-end); but if a large magnet needs to be scaled up with the magnetic field kept constant, the volume, weight, and hence the major part of the cost will go up as the cube of the radius. But the energy of particles in their equilibrium orbits goes up only as the radius or, if they are still moving much more slowly than light, as the square of the radius. Thus for sufficiently high energy a linear accelerator will be cheaper than a scaled-up magnetic device. The idea is an old and unsuccessful one; but the familiarity with microwave techniques gained in wartime makes it appear practical once more. Professor Alvarez at Berkeley is now engaged in building a large linear accelerator for protons. Protons enter the machine with a few million volts energy, from a one-step accelerator, in this case a Van de Graaff machine. The vacuum tube down which they fly is one long resonator, made of many resonant cavities placed back-to-back and so driven by individual but phased microwave oscillators at 150 cm that the moving protons enter each cavity in phase with an accelerating electric field. By the time they cross the cavity, one cycle has elapsed, the field behind has dropped to zero, and the next cavity is beginning to acquire an accelerating electric field. There is one difficulty with this simple picture. The protons do not move with uniform velocity, but constantly increase in velocity. If the cavities are driven, as they must be, by an oscillation of one frequency, some trick must be used to keep the protons in step. This is done by adjusting the repeat length of the resonant cavities $l$ so that $l/\sigma = \lambda/c$ where $\lambda$ is the wave-length of the driving oscillator and $\sigma/c$ the velocity of the particle relative to that of light. If the outside cavity diameter is held fixed, the cavities must be modified in shape to keep them resonant, though of different lengths. In transmission line language the line is a loaded line. A short coaxial cylinder is placed inside the outer cavity cylinder, and the protons move down the axis of the whole cavity. Adjusting the diameter of these central tubes keeps resonance. Each cavity will during half the cycle be decelerating: the electric field will point the wrong way. During this time the proton must be inside the central tubes, and shielded from the wrongly-directed field. The tubes are hence called drift tubes, because within them the protons "drift" under no field.[13] To maintain phase stability and geometrical focusing, it is necessary to place conductors athwart the drift tubes, through which the protons must pass. Thin beryllium foils have been developed for this work. In trials more than a million volts per foot has been developed in a cavity. As yet no coupled cavities have been tried for acceleration, but the first forty-foot unit is soon to be tried out. It is expected that it will be delivering protons at more than thirty-two million volts by February, 1947. New high power radar oscillators have been developed for the job, using the pulsed magnetron principle. Here again the output beam will be pulsed at a rather slow audiofrequency, for the sake of magnetron cooling and performance.

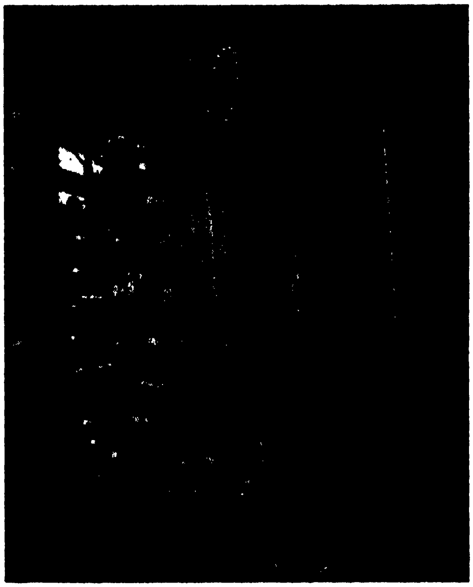


FIG. 10. A close-up of the rotating vacuum condenser whose housing is seen in Fig. 9. The many teeth which produce the varying capacity are plainly visible. The oscillator frequency, which is in the neighborhood of 10 Mc, is varied about 20 percent at a modulation rate of a few hundred times per second by means of this device.

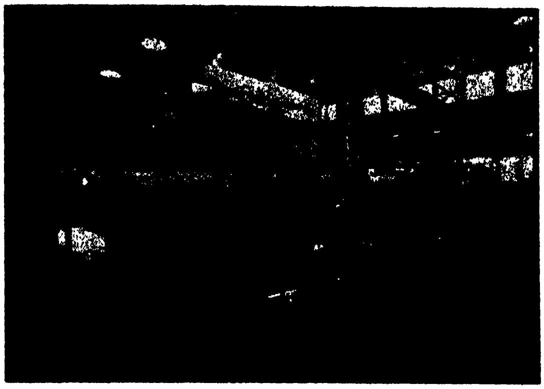


FIG. 11. The forty-foot linear accelerator at Berkeley. Note the large tank of the horizontal pressure Van de Graaff machine at the far right. This is the ion source for the machine. The outer steel tank for the linear accelerator tube itself is conspicuous, flanked by the thirty-odd war surplus 1.5-m radar oscillators on either side. Protons in the thirty-to-forty-million volt range are expected by February, 1947. New high power magnetrons are under construction to replace the many oscillators shown here. Only a few of the new magnetrons would be needed for a forty-foot section. (Photo from Professor L. W. Alvarez, heading the linear accelerator group.)

A similar linear accelerator but without the drift tubes and without phase stability can be built for particles moving near the speed of light.[14] High energy electrons are planned for at M.I.T. and at Purdue by groups working on the design of such a device.

If mesons are not made in 1947, it will not be for want of effort! It is to be hoped that enough leisure will be left to plan the experiments which these machines will make possible.

## THE LEGACY OF THE WAR

The year was marked by the widespread if delayed publication of results of the key war projects in physics, and above all, by the return to their old laboratories of hundreds of war-experienced physicists, brimful of information about what had been done, and confident in their understanding of whole fields of technique which had been vague general possibilities in 1940.

---

[13] Oppenheimer, Johnston, and Richman, Phys. Rev. 70, 447 A (1946); L Alvarez, Phys. Rev. 70, 799 A (1946).

[14] J. Slater, Phys. Rev. 70, 749 A (1946).

### 1. Radar

To this author one of the most fascinating parts of microwave radar has been the duplication in the radio spectrum of the familiar results of physical optics. The "optics" of microwaves—mirrors, gratings, diffraction at openings—are by now familiar to everyone. Not so familiar is the microwave lens.[15] An optical lens works, of course, because the phase velocity of the electromagnetic wave is reduced by the coherent scattering of the bound electrons in the dielectric. Shaping the lens is a matter of adjusting the delay introduced along each ray path to produce a net wave front with the form desired, usually changing the radius of curvature while leaving a spherical incoming wave still approximately spherical. No one has made bound electrical oscillators on a larger scale for microwave optical glass. But the fact has been used that the propagation of a wave in a wave guide, a region bounded by conducting walls, proceeds with an increased phase velocity, *greater* than that of light in free space. (If the reader is worried about relativity, he should review the distinction between wave and group velocity.) By spacing a set of copper plates, their planes parallel to the direction of propagation of the wave, a half-wave-length apart, a change in the phase velocity of the microwave can be secured. Shaping these plates in forms not unlike those of ordinary optics (but with a refractive index always less than one) microwave lenses can be made which quite successfully act as condensing lens, and even produce a fairly sharp focus. They are of course good only for one wave-length. "Chromatic aberration" is large!

Quite the opposite point of view has been used with good effect in recently-reported experiments using microwave radar technique. These experiments were really experiments in measuring the absorption spectrum of water vapor molecules in the region of wave-lengths from 7 mm to about 1.7 cm. An absorption line was found at 1.34 cm and its shape measured as a function of the pressure. The same line could be predicted roughly from the older work on the rotational fine structure in the infra-red region around 20 microns wave-length. Absorption is of course a quantum phenomenon, and it is satisfying to see that the analysis of the instrument—the "spectroscope"—employed came very naturally if one remembered the quantum nature of microwaves.[16]

The instrument consisted of an eight-foot cubical box of copper, arranged to be filled with moist air. Through the box were strung, at random, the detectors, in this case thermocouples. The box was coupled to a pulsed magnetron oscillator—several were used to get points at several wave-lengths—and the space-and time-averaged energy density measured as proportional to the thermoelectric e.m.f. This energy density for a constant excitation is compared with and without the water vapor. The oscillator pumps a fixed number of quanta into the box each second. These quanta are eventually absorbed, either by the walls of the box, the thermocouples, solder, air, and fingerprints of any actual experiment, or by the molecules of water vapor in the box. Since the thermocouple reading is proportional to the density of quanta built up in the box, all that is required is to compare the rate of loss due to the water vapor with some known absolute rate of loss of quanta, and the absorption of water vapor is known for that particular wave-length. But how to get an absolute loss rate? It is difficult to compute such a quantity for walls of a copper cube, the more because of the actual complex nature of the experimental geometry. But the quantum idea gives a direct method. It is necessary only to make a hole in the side of the box. All the quanta that strike the hole must leave the box. This number can be calculated just as the number of molecules leaving an orifice to form a molecular beam can be calculated. The expression is just the familiar kinetic theory formula: the number of quanta leaving through a hole of area $A$ per second is given by $\frac{1}{4}nc$, where $n$ is the number of quanta per unit volume and $c$, of course, their mean velocity—the velocity of light. The whole method of procedure was checked both by experiment and by the more complicated exact calculations of wave theory.

Randomness had to be guaranteed. No standing wave pattern could be allowed. This was as-

---

[15] W. E. Kock, Proc. I. R. E. 34, 828 (1946).

[16] W. E. Lamb, Phys. Rev. 70, 309 (1946); Becker and Autler, Phys. Rev. 70, 300 (1946).

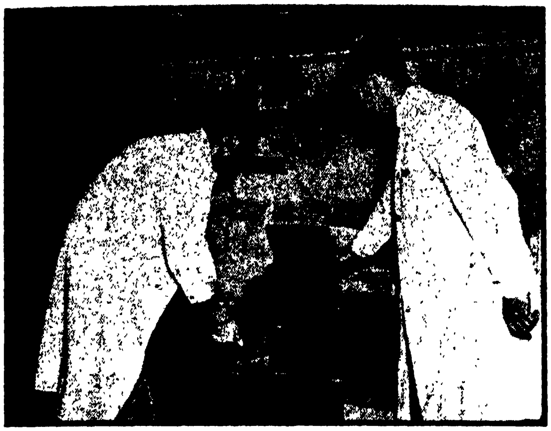


FIG. 12. A face of the shielding wall of the graphite-moderated pile at the Clinton Laboratories at Oak Ridge. The operation shown is that of exposing a sample to a strong neutron flux, for the production of a radio isotope. The sample in an aluminum holder is being inserted into a hole in a graphite block can be pushed back along a slot into the interior of the pile. The slot is then filled with more graphite blocks to complete the structure, and the shield hole is then plugged. Note the many removable "stringers" of graphite for which shield openings are provided.

FIG. 13. A face of the "hot lab" at the Clinton Laboratories. This is a concrete cubicle fitted to permit the remote operation of chemical plant adequate for the inorganic separations needed to produce pure radio isotopes from fission product. The air and steam valves and lines, the thermocouple potentiometer, plainly seen in the right foreground are typical for any chemical operation. The periscopes through which the operators are watching the reaction behind the concrete wall are the evidence of the special problems met in the handling of kilogram equivalents of radium.

sured by the slow rotation of some copper-bladed fans within the enclosure, which changed the geometry and made it sure that a random set of normal modes was excited. Diffraction effects were small, because the hole was made large compared to the wave-length. Apertures up to several hundred square centimeters in area were used. The whole experiment not only gave results and pioneered a method sure to be useful for the study of molecular structure, but formed a beautiful illustration of the unity of our dual picture of radiation.

A number of other gas absorption studies have been made at microwave frequencies, using more conventional techniques, like attenuation in a wave guide, or using the atmospheric propagation itself.[17]

## 2. Atomic Energy

The chain reaction of uranium and its consequences are too well known and too extensive to give any general account here. The slow but steady publication of the results of this project will eventually form a sizeable branch of nuclear physics, as they now do in secrecy. The great continuing laboratories at Chicago, Oak Ridge, Los Alamos, and Berkeley, as well as the new regional laboratory at Brookhaven, Long Island, and the power development laboratory at Schenectady will become an integral part of American physics. Some topics selected out of the many possibilities from the Manhattan project will be sketched here.

### *a. Isotopes*

Some hundreds of separate lots of artificial radioactive isotopes have been sold to research workers already by the Isotope Section, Research Division, Manhattan Engineer District, PO Box E, Oak Ridge, Tennessee. This activity will surely continue and expand under the newly formed Atomic Energy Commission, which took over control of the great project from the Army on January 1, 1947. The long-lived soft beta-emitter, $C^{14}$, made in the high neutron flux of the Clinton graphite pile by the reaction $N^{14}(n,p)C^{14}$ which goes well with thermal neutrons, is the material of greatest interest as a tracer in biological studies. The two-week beta-emitter, $P^{32}$, is also in demand, chiefly for its therapeutic value in some cases of leukemia and related diseases. It is made by neutron capture in the

---

[17] Beringer, Phys. Rev. 70, 53 (1946); Dicke *et al.*, Phys. Rev. 70, 340 (1946); Townes, Phys. Rev. 70, 665 (1946).

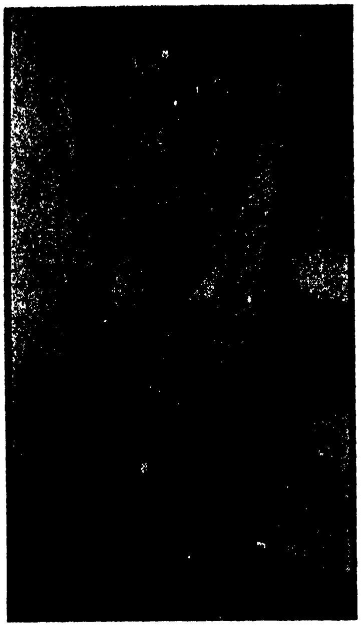


FIG. 14. A microphotograph of the first sample of the new element americium ($Z$-95) ever isolated. The lower part of the photograph shows the eye of an ordinary needle to fix scale. Above it is a small glass capillary, which is the test tube for this sort of chemistry. In the capillary can be seen the curved meniscus of the supernatant solution. At the bottom tip of the *inside* of the glass capillary is the grayish precipitate of americium hydroxide. (Photo furnished by Professor G. T. Seaborg, discoverer of Am.)

normal phosphorus isotope, $P^{31}$. A third popular product is $I^{131}$, which is a fission product, produced in the exposed uranium metal slugs of the pile. These three are only examples, but they indicate the three types of reactions which go with ease in the reacting pile. The technique of separating a particular radioactive isotope out of the complicated mixture which may be formed by bombardment is well shown in Figs. 12 and 13. It always involves remotely controlled chemical operations, with all personnel protected from the radiation emitted by thousands or tens of thousands of curies of activity. The handling of large amounts of radioactive substances is now a well-developed branch of engineering.

### *b. The Transuranic Elements*

The element of highest atomic number and highest atomic weight which is found in nature is of course uranium. But the periodic table now goes well beyond that. In the year 1946 the disclosure and naming of the transuranics had progressed through neptunium, number 93, and plutonium, 94, to americium, 95, and curium, 96.[18] In 1942 plutonium was produced only by cyclotron bombardment beginning with natural uranium. It was obtained in microgram amounts, and the extraordinary techniques developed by the radio chemists to handle such ultra-microchemistry have been admired widely by now. The elements at the top of the table are still made in microgram amounts by cyclotron bombardment. But the starting point may now be the man-made element plutonium, available in kilogram lots! It is interesting that the chemistry of the transuranics, especially of trivalent curium and americium, seems to prove that a new rare earth like series of the periodic table begins with actinium, atomic number 90, and that successive elements essentially are made by adding electrons, not to the outermost $7s$ shell, but to fill the $5f$ shell. This is in close analogy to the rare earths, which may be called the lanthanides, for at lanthanum the $4f$ shell begins to be filled. The transuranics we should call actinides.

It is appropriate here to say that identity and name are now suggested, though not yet officially, for all the elements of the table up to curium, 96.[18] The blanks are to be filled in this way, all by known radioactive isotopes of the element in question: Element 43 will be named by Professor Segrè of Berkeley, probably to be called technesium, from *techne-* art, to indicate that it is artificial only. It has no stable isotopes. The previous identification of a stable 43, called masurium, is certainly in error. Element 61 is now known to be a rare earth fission product. No name has been suggested as yet. Element 85, an unstable halogen, will be called by the beautiful name astasine, from *a-* not, and *stasis-* standing still. Element 87 is known from radioactive work only, not chemically, and on these not quite complete grounds may be named francium, since it was discovered in this sense by a French radio chemist, and is believed to be a homologue of cesium.[19]

---

[18] Seaborg, Science (Oct. 25, 1946).

[19] See, for example: Seaborg and Segrè, Phys. Rev. 55, 808 (1939) for element 43, Corson, MacKenzie, and Segrè, Phys. Rev. 57, 459, 1087 (1940) for element 85, Perey, J. de phys. et rad. 10, 453 (1939) for element 87. Private communications from E. Segrè, D. Corson, C. Coryell, and A. Turkevitch.

### *c. The Pile as an Instrument*

The chain-reacting pile forms an excellent device for the determination of the probability of neutron absorption, especially for thermal neutron absorption, by any material. The substance is placed into the reacting structure at a point where the neutron density is fairly high. At such a place, evidently, the absorption of neutrons will have an important effect on the criticality of the whole structure. Since the time required for the neutron density to double in a barely-critical pile is approximately proportional to the difference between the actual reproduction factor of the structure and unity, careful measurements of pile level changes with and without samples introduced provide a very sensitive way of measuring neutron absorption. Macroscopic samples are used, of course, and the cross section per atom can be measured at least relatively with high accuracy. This method has been applied to many elements—all that could be tried—and even, for engineering purposes, to manufactured materials. By oscillating the sample to and fro, from one point of the pile to another where the neutron density was quite different, sinusoidal change in neutron level of the whole pile is produced.[20] This allows the elimination of slow drifts and the amplification of the effects by the use of resonant galvanometer systems. In the hands of the group at Oak Ridge this has become a method suitable for both precision measurement of absorption cross sections, in favorable cases, and in the detection of very small cross sections. A general review of all the neutron cross sections over the whole periodic table is in preparation based on the extensive project work. The pile work was mainly done at the Argonne Laboratory in Chicago and the Clinton Laboratory at Oak Ridge; very beautiful work in the region of fast neutrons was done mainly at Los Alamos; and neutron spectroscopy, with cyclotron velocity selectors or with crystals (see below) at all these places and by sub-contract at Cornell and Columbia. Much of this work is in the press for release; some has appeared this past year.[21]

### *d. The Pile as a Source of Neutrons*

A well-collimated beam of thermal neutrons defined by cadmium slits only millimeters wide and meters apart was long a dream of neutron physicists. When the intensity in addition leaves little to be desired, real progress has been made. This was made clear in 1946 by the experiments reported from the Manhattan Project—still only fragmentarily—in which the beam of neutrons from a pile was put to work. The Bragg crystal diffraction of neutrons is not a new effect. It has been somewhat obscurely demonstrated in several laboratories. A neutron moving with the energy corresponding to thermal equilibrium in the lattice of the block of graphite in which it has been slowed has a wave-length of a few angstroms. This is just right to give strong low order diffraction maxima from natural crystal gratings. The effect has been put to work in the construction of crystal spectrometers,[22] using not x-ray tubes, sources of the diffracted waves, brass slits, and photographic plates for detectors, but cadmium slits defining a sharp beam of thermal neutrons from the pile, and boron-filed ionization chambers as detectors. The apparatus has had many uses. At Chicago[22] the neutrons have been used to investigate the scattering of neutrons from crystalline compounds. The neutron is scattered with different phase and amplitude by different nuclei. The intensity of scattering

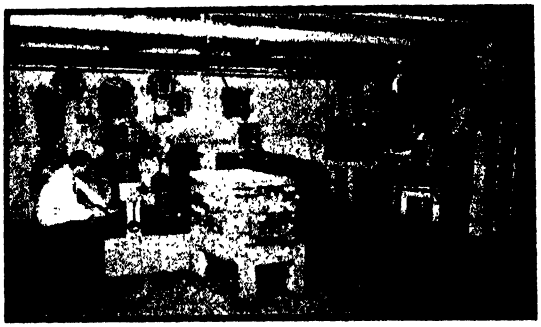


FIG. 15. Another section of the face of the Clinton pile. This suggests at least the kind of geometry applicable to the production of a strong collimated neutron beam. A beam is here emerging from a small hole in the pile face (the opening marked 20) and being caught in the large lead brick housing in the center of the photo.

---

[20] E. O. Wollan, in press.

[21] For example: Bailey *et al.*, Phys. Rev. 70, 583 (1946); Barschall, Battat, and Bright, Phys. Rev. 70, 458 (1946); Rainwater and Havens, Phys. Rev. 70, 136, 154 (1946); H. H. Goldsmith, a survey article in press.

[22] Abstracts by Fermi, Zinn, Sturm, Turkel, and L. Marshall, Phys. Rev. 70, 103 (1946); Borst *et al.*, Phys. Rev. 70, 557 (1946).

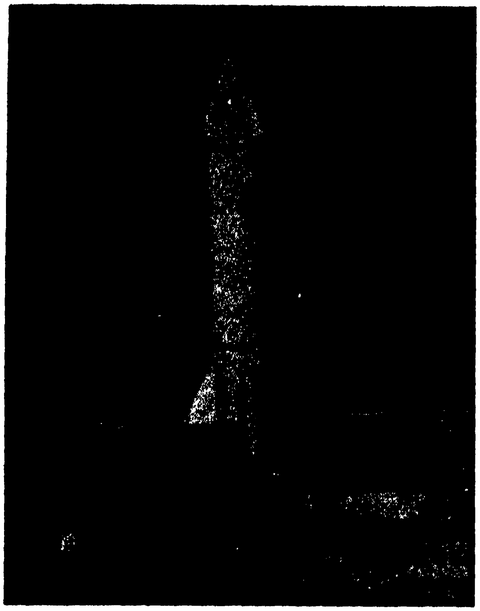


FIG. 16. A German-built V-2 rocket being made ready for flight at the proving grounds in the White Sands region of New Mexico. Note the truck at the left evidently supplying the liquid oxygen fuel to the rocket.

determines the probability of the process, or the scattering cross section. This has long been investigated. But the interference of the scattered amplitudes from the planes of differing nuclei in crystals, or even in molecules, gives a way to measure the hitherto unmeasured phase shifts in the scattered neutron wave. They show up in the relative intensities of the different orders of diffraction from various crystal planes, for example. This effect has no parallel in x-ray diffraction, where the wave is electromagnetic and its phase shift uniform. When this complication is unravelled, use of the neutron diffraction as a supplement to x-ray crystal analysis is likely to interest the workers in this field. Neutrons feel the effect of hydrogen and especially deuterium atoms much more than do x-rays, which are capable of interaction only with the electron cloud. This may lead to the study of the hydrogen atom location in some compounds with the aid of neutrons. The neutrons which have been thermalized leak out of the pile with a velocity distribution just that of a molecular beam of hydrogen leaking out of a chamber at a little above room temperature. Collimated and directed suitably against a crystal of salt or calcite or what you will, a spectrum of the neutrons will be spread out in angle, according to the well-known Bragg law, $n\lambda = d\sin\theta$. Here $\lambda$ is the deBroglie wave-length, $\lambda = h/mv$, and thus the distribution in angle is a distribution in velocity. Putting absorbers in the diffracted beam at any angle will serve to produce the attenuation for the energy of neutrons present at that angle of deflection. This is a physically monochromatic beam of neutrons, not simply a device for selecting the effects of a particular velocity group, like all the familiar time-modulated schemes. Energy resolution and range of usable energies is about as good as the best time-modulation schemes, and—apart from the pile—very much simpler. This scheme has been responsible already for the mapping of several resonance peaks in neutron absorption, for example those of In, Rh, Sm, Eu, and Gd.

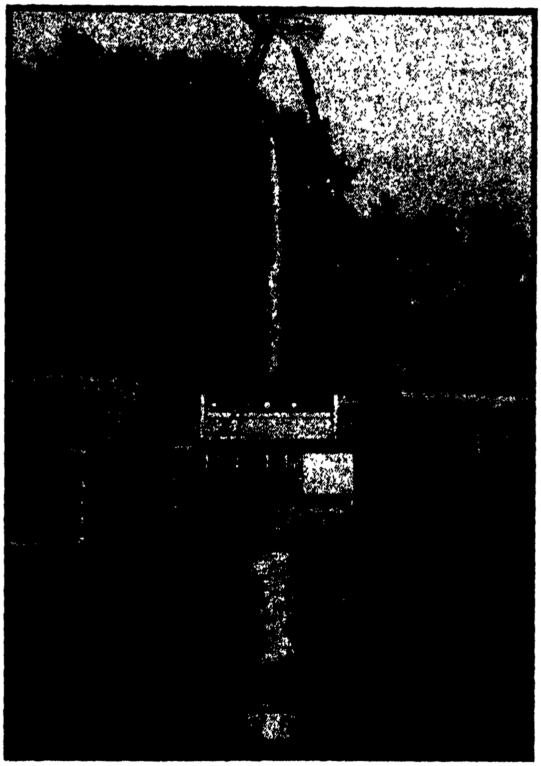


FIG. 17. A war-head of the V-2 fitted with research apparatus, in this case the electronic circuits of a cosmic-ray telescope whose Geiger-Muller tubes "look" through the openings in the war-head. This installation was made by the Applied Physics Laboratory of Johns Hopkins, at Silver Springs, Maryland.

The strong beam has given us one result of decisive importance, a precision re-measurement of the magnetic moment of the neutron.[23] At the Argonne Laboratory, the heavy-water moderated pile, which produces a strong thermal beam, was used as the source for an experiment measuring the magnetic moment of the neutron by polarizing and analyzing the neutrons with ferromagnets through which they passed. The magnetic moment was obtained by observing the resonant frequency at which the polarized beam was partly depolarized by induced transitions in a magnetic field, coming from the Larmor precession of the neutron magnetic moment. (This is the exact experiment of Alvarez and Bloch, an analog to those of Rabi with molecular and atomic beams.) The proton and deuteron moments were measured as calibrating data in the same magnetic field, using the method of nuclear induction in one form (see below). The result was given a conservatively estimated precision of one part in fifteen hundred. To this accuracy the magnetic moment of the deutreron is shown now to be just that of the free neutron plus that of the free proton plus a small calculable contribution arising from the non-spherically symmetric motion of the proton in the deuteron, the known slightly cigar-shaped distribution of charge. No effect of the binding on the nuclear particles themselves is found, to a rather high accuracy. This point is a long debated one in the theory of the lightest nuclei, and sets a necessary condition to be met by any meson theory of the future.

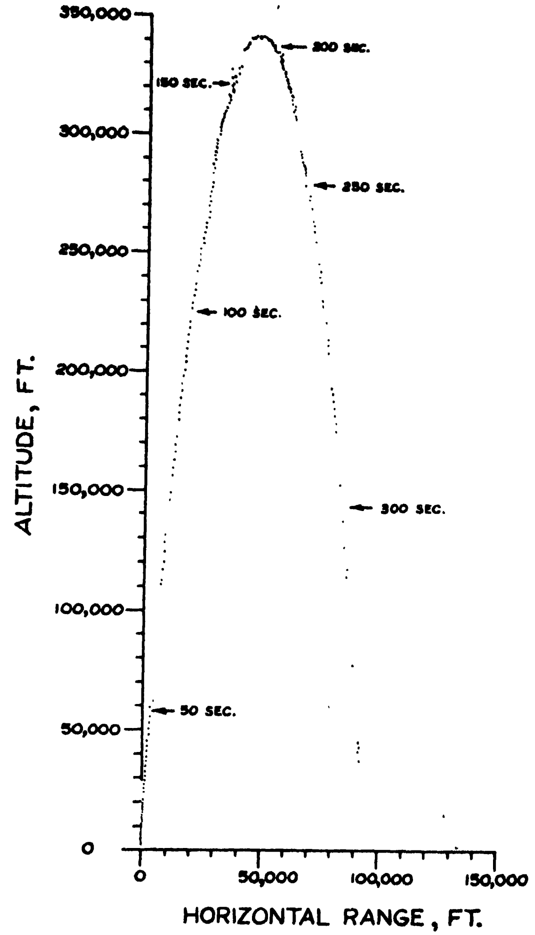


FIG. 18. The graph of the spectacular trajectory of a successful V-2 flight. The high altitude data were obtained by radar tracking. Note that about four minutes is passed in regions beyond the reach of balloons or planes, above 100,000 feet. Not long, but something!

## 3. Rockets

The great technical development of the Axis Powers during the war was certainly the V-2 rocket. For the first time it was possible to launch a projectile of some size into the region beyond the atmosphere. This wartime feat again has meaning for physics. Last year two dozen rockets were launched in New Mexico, most of them reaching altitudes of fifty miles or more. The flight lasts for several minutes, and up to a ton of apparatus can be carried aloft. In cooperation with the ordnance and industrial teams engaged in studying the rockets as weapons, several laboratories have taken advantage of the chance to study the region beyond the atmosphere. Here are the cosmic-ray primary particles, not yet complicated by cloudbursts of secondary particles which they cause on striking the atmosphere. Here may be measured the spectrum of the sun and of the stars, not through the dark glass of the air, but as they come through empty space. Both spectra and cosmic-ray measurements have been made with interesting results.[24] Too little has yet been done to draw any valid conclusions, but it is certain that the exploration

---

[23] Arnold and Roberts, Phys. Rev. 70, 766 (1946); cf. Alvarez and Bloch, Phys. Rev. 57, 111 (1940).

[24] Golian, Krame, and Perlow, Phys. Rev. 70, 776 (1946); private communication from Applied Physics Laboratory, Silver Springs, Maryland.

of space beyond the air will have real importance for more than one field of physics.

### THE MASS OF THE MESON

For the years since its discovery, the meson has been examined in the cosmic ray in an effort to measure all its properties. Most fundamental, after its charge, is perhaps its rest mass. If the cosmic-ray meson is that predicted in connection with nuclear force theories, its mass should be somewhere between 150 and 250 times that of the electron. Up until 1946 about twenty-five mesons had been observed with sufficient accuracy to make possible some measurement of their mass. The spread in values was great, and it was by no means sure that the meson had only one definite mass. Perhaps the meson was not a unique particle but a whole family.

An excellent experiment just reported[25] seems to have shown that the ordinary meson which makes up the penetrating component of cosmic rays at sea level has one mass, about 202 ± 5 times the electron rest mass.

This measurement was made by a very careful and successful application of the usual technique. The method of measuring the mass is ordinarily this: The momentum of the meson can be measured by measuring the curvature of its path as made visible in a cloud chamber across which there is applied a known constant magnetic field. This measurement is made difficult by the accidental deflections of the track of droplets caused by the irregular and turbulent motion of air in the cloud chamber. Then the ionization which the particle produces is measured, by counting the droplets left in the chamber gas, or by seeing what thickness of matter it will penetrate before it comes to rest. This gives the rate at which the particle loses energy by interaction with the electrons of matter, and this is known as a function of the velocity of the particle. The scheme then is to measure the curvature of the track and on the same particle to determine what thickness of lead it will penetrate. The experiment here reported divided up these measurements between two cloud chambers. The upper chamber was free of any obstacles, and carefully controlled in temperature. Track curvature could be measured with minimum difficulty from turbulence. A second chamber mounted two feet below the first and in the same plane was expanded simultaneously, and photos were taken of both chambers. The lower chamber was crossed by eight half-inch sheets of lead. By observing in which plate the track appeared to end the range in lead could be gotten to fair accuracy, and the velocity calculated. The particles observed could be assigned a momentum with a spread due to the error in curvature measurement, and a range with the error coming from the finite plate thickness and other sources. The particles then could each be assigned a mass and a definite error. The spread of values was from 142 to 264 electron masses. Statistical analysis of the data gives the result that a unique mass of 202 *m* is quite consistent with the data. The observations agree among themselves just as well as can be expected from their individual estimated errors. These mesons have the single mass 202 *m*.

### NUCLEAR INDUCTION

For some years the magnetic moment which is intrinsic to nuclear particles and to their motion within the nucleus has been an important object of study. This is a very small magnetic moment indeed, roughly two thousand times less than that associated with a single atom. First evidence for and some measurements of these tiny magnetic dipoles were spectroscopic. They interacted with the current respesented by the orbital motion of the electrons around the nucleus in the atom, and the different orientations the nuclear magnetic moment assumed in the magnetic field resulting from the electrons' motion gave rise to atomic energy levels. These levels are very close together: the energy difference is very small, and the difference in frequency and hence wavelength of the emitted spectral lines very small indeed. Lines which originate in this way are said to be lines of the hyperfine structure of spectra. About ten years ago another and more elegant method was devised, which reached its present form in the work of Professor Rabi and his co-workers at Columbia. Here the nuclei are examined as they stream in molecular beams. Such beams are made to pass through strong inhomogeneous magnetic fields. The nuclear mag-

---

[25] W. B. Fretter, Phys. Rev. 70, 625 (1946).

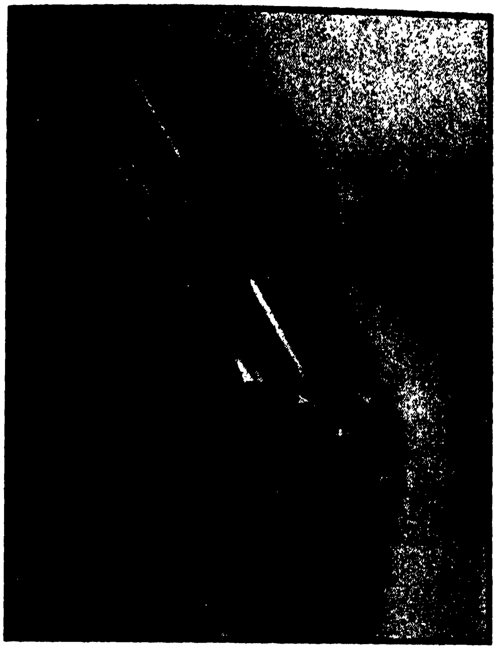


FIG. 19. The receiver-transmitter coils first used to measure the nuclear magnetic moment by nuclear induction. The four heavy turns are the transmitter coil, and the more finely-wound receiver coil is seen within, its axis at right angles to the transmitter coil. Within the receiver coil is placed the sample encased in a half-inch spherical glass bulb. The whole arrangement is shielded heavily. Openings in the shield allow the installation of a rotating copper paddle for adjusting stray fields from without. (Photograph by Professor Felix Bloch of Stanford University, where this work was done.)

netic dipoles feel a force which will depend upon their quantized orientation in space, and the molecular beam will split into several components. These components are each deflected by a different amount, and the deflection may be measured. The deflections one gets with any realizable magnetic fields are small, of the order of hundredths of a millimeter. The method could not by the nature of things be very precise. An elegant adaptation of the molecular beam technique was made by Rabi. He used two equal but oppositely directed inhomogeneous fields. The beam passes through the first and is split into its components. Then the beam is reunited by passage through the second, and strikes a detector. But between the analyzer and the second field there is placed a third region. Here a uniform magnetic field is produced, and also a small radiofrequency magnetic field. The large uniform field has of course no effect on the nuclear

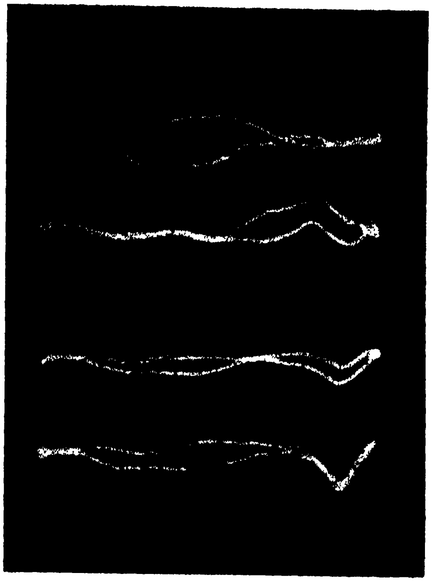


FIG. 20. Oscilloscope traces of the nuclear induction signal. The vertical deflection is proportional to the precessing magnetic moment (or its component in the direction of the axis of the receiver coil), the horizontal deflection to the applied d.c. magnetic field, which has a small, 60-cycle variation in magnitude. The traces shown are those of the proton signal from a water sample. In the top trace the applied field is above the resonance field on the average. The d.c field was lowered in about a second's time to a value below resonance. The proton signal decreased slowly. The next three traces were taken at successive five-second intervals after the reduction of the d.c. field. Note that the signal slowly reverses to a below-resonance condition, though no external change is now going on. This time lag, or "memory," is the consequence of the time it takes for the small nuclear magnets to realign themselves into thermal equilibrium with the new applied field. The time is clearly a few seconds in the case of water. (The double trace is a result of stray 60-cycle pick-up, which separated the forth- and back sweep in the camera exposure.)

dipoles. But the radiofrequency "quanta" may be absorbed whenever their frequency reaches a value such that the quantum energy $h\nu$ equals the energy difference between one orientation of the nuclear magnet in the uniform field and another one. Classically one writes $h\nu = \mu H$, where $H$ is the uniform field strength and $\mu$ the magnetic moment of the nucleus. Fields around a thousand gauss correspond to resonant frequencies of a few megacycles for typical nuclei. Now the nucleus which "absorbs" the quantum is flopped over to a new orientation, cannot be refocused in the second field, and never reaches the detector. By measuring the detector response as frequency or magnetic field are varied, the shapes of the lines corresponding to resonance *with the nuclear magnetic moment can be found.* This leads to precision comparisons of nuclear moments, depending only upon frequency measurements, always easy to do well. No nuclear quantities are known with the precision with which nuclear magnetic moments can be found in this way. But the apparatus is delicate and the *skill and patience required great.* Nor can every substance be gotten into a molecular beam and detected.

It was exciting, then, when not only one but two investigators suggested methods by which the same phenomenon—the flip of the nuclear magnets when they feel an oscillating field of the right frequency—could be detected with no fancy molecular beam techniques. For the first time the macroscopic effect of the nuclear magnets on the magnetic polarization of bulk material was taken into account. Think of a sample of material placed in a magnetic field. The nuclear magnets will line up in the magnetic field to get into the state of minimum potential energy. But of course the thermal agitation of the molecules will act to disorient the magnets. The net result will be given by the usual Curie-Langevin law: the resultant nuclear moment per cc at thermal equilibrium will be approximately $n\mu(\mu H/kT)$, where $n$ is the number of nuclei per cc with magnetic moment $\mu$ per nucleus, $H$ is the applied magnetic field, and $kT$ is Boltzmann's constant times temperature. This is a small contribution to the paramagnetism at best. For protons in water in even pretty high fields it is almost unobservable. What is measured is not the d.c. effect, but a resonant effect at radiofrequency. In one experiment, a resonant cavity was made for about 30 mc.[26] It was filled with paraffin, and was placed in a strong magnetic field, at about 7000 gauss. A weak r-f magnetic field in the cavity was kept perpendicular to the d.c. field. The d.c. field was slowly varied until a sharp resonance absorption was observed. The absorption of r-f energy had changed the $Q$ of the cavity, reduced its output, and affected a detector which had been balanced out off resonance. The same advantages can be claimed for this method as for the molecular beam.

An even simpler technique was applied somewhat later. Here the effect observed is the precession of the resultant nuclear magnetic moment as resonance is approached.[27] The observation is made, not of the reaction upon the driving circuit, but of the e.m.f. induced directly by the precessing nuclear magnets in a pick-up coil in which all e.m f. had been balanced out off resonance. The nuclear-induced precessing field has a component perpendicular both to the constant field and to the r-f field. It is this component which is observed. No more is needed than a radio oscillator, a lecture table magnet, a transmitter-receiver coil arrangement, and appropriate radiofrequency amplifiers and oscillograph circuits. The coils of Fig. 19 are the heart of the apparatus.

It should be noted that the methods of nuclear induction, as they are called, depend upon the nuclear magnets reaching thermal equilibrium, with their moments oriented not at random, but with the equilibrium resultant value. It is just the macroscopic want of cancellation which makes the whole effect. But the mechanism by which the nuclei come to thermal equilibrium is complex and little-known. If it took weeks for equilibrium to arise, the experiment would be very difficult. The time taken will be a function of the chemistry of the compound and of many atomic features. A whole new subject in atomic physics is opened up by this technique, as well as a simple new supplement to the existing study of magnetic moments. The use of the technique for isotope analysis without any destruction of the sample, isotope analysis by radio, so to speak, may prove of great importance in tracer work with stable isotopes. The whole subject is a good example of how new ideas may arise in fields believed already carried to their highest development. The world of physics is surely infinite.

This sketchy review of the first year of peace is full of promise. The promise will be fulfilled only if physicists can share with all men the prospect which carries all our hopes, the prospect of the many years of peace that lie ahead.

I am glad to acknowledge the kind cooperation of all the busy men who answered letters and supplied photographs to make this account possible.

---

[26] Purcell, Torrey, and Pound, Phys. Rev. 69, 37 (1946).

[27] F. Bloch, Phys. Rev. 70, 460 (1946) (theory); Bloch, Hansen, and Packard, Phys. Rev. 70, 474 (1946) (exp't).

# Contributed Original Research

## An Effect of Electron Bombardment upon Carbon Black

JOHN H. L. WATSON
*Plant Research Department, Shawinigan Chemicals Limited, Shawinigan Falls, Quebec*

(Received September 9, 1946)

A significant change is reported in mean particle size and shape characteristics of carbon black due to specimen contamination while under examination in electron microscopes. The effect is described for a number of well-known commercial blacks and examples are given in graphical form to illustrate expected variations in mean particle size with continued bombardment at normal focusing intensities. The effect is observed in other materials but is more pronounced in carbon black. Suggestions are made for minimizing the effect.

A CHANGE in any specimen in an electron microscope resulting from the effect of either vacuum or electron bombardment is of interest to electron microscopists both for itself and for the information which may be gained for interpretation of electron micrographs. Very few effects of vacuum upon electron microscope specimens have been reported, although an evaporation of sulphur crystals has been observed. Some modification of specimens in the electron microscope due to shrinkage or evaporation, is discussed in a recent article by L. Marton *et al.*[1] Artifacts introduced into electron images by electron bombardment of a specimen object are also rare. Dulling of sharp crystal corners and edges by bombardment has been noted and charging effects are common, but permanent deformations are not often seen. One notable exception involves the phenomenon of the deposition of insulating films upon metal surfaces under electron beams.[2]

The particular effect reported here is an increase in the mean size of carbon black with bombardment as measured by the methods of electron microscopy. This was observed first with a filmless[3] preparation where the increase is very pronounced. This increase in mean particle size is important from the point of view of its explanation and from the practical standpoint of its influence upon the problem of carbon black particle size and shape determinations.

A qualitative description of the change in appearance of Shawinigan Acetylene Black under bombardment can be described best from an examination of Fig. 1. This figure shows the appearance of a particular field of a filmless sample as the time of bombardment increases. Bombardment times for each exposure are given in the figure caption and refer to the total elapsed period relative to the first exposure. As the bombardment continues the original crystalline appearance of the particles is lost and almost all become round or oval after a few minutes, where previously they were predominantly crystalline and straight-edged. The particles not only change shape, but they increase in size, and fewer are distinguishable. There is a lessening of contrast over the image, less structure is visible in both chains and particles, and the chains become clumps. The change in mean particle size and in appearance is considerable over even short periods of 15 or 30 seconds, but occurs slowly enough that the eye can distinguish it only after discrete intervals. Over the usual times allowed for examination of a specimen in the electron microscope the changes may be unnoticed qualitatively but will be of sufficient magnitude to materially affect particle size frequency distributions and observations of shape characteristics.

The effect has been checked for many preparations of Shawinigan Acetylene Black and has been observed to some extent in all samples studied.

---

[1] L. Marton, N. N. Das Gupta, and C. Marton, Science 104, 35–36 (1946).

[2] R. L. Stewart, Phys. Rev. 45, 488–490 (1934).

[3] J. H. L. Watson, J. App. Phys. 17, 121–127 (1946).

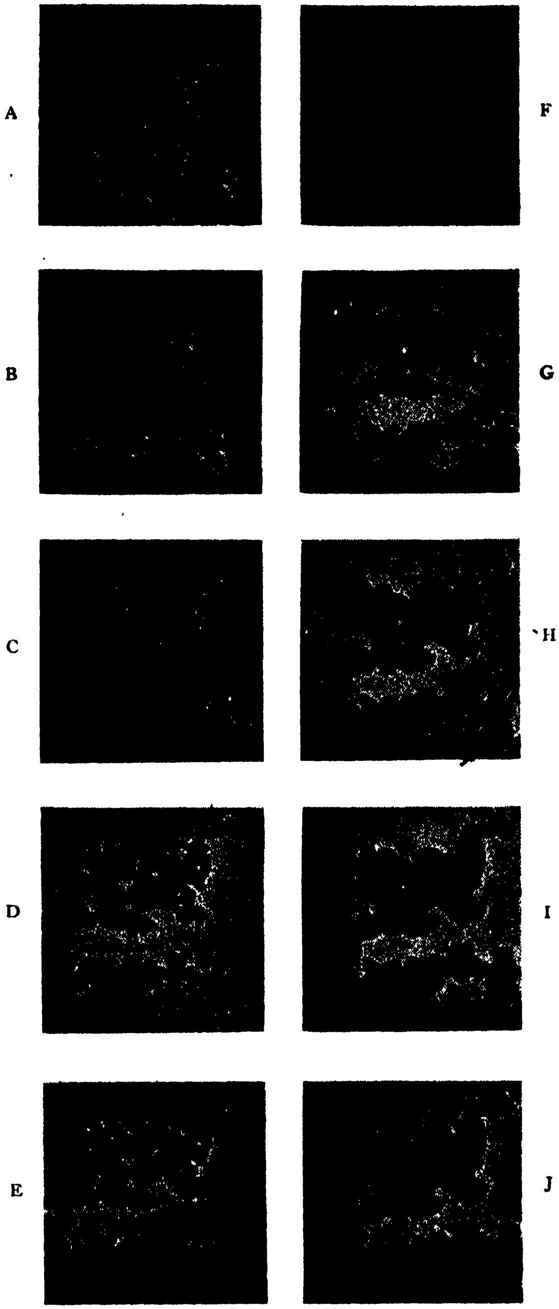


FIG. 1. Shawinigan Acetylene Black under continued bombardment in the electron microscope, filmless preparation ×20,000. Bombardment times are: A, zero; B, 1 min.; C, 2 min.; D, 3 min.; E, 4 min.; F, 5 min.; G, 7 min.; H, 9 min.; I, 11 min.; J, 12 min. relative to the first exposure.

Figure 1 is representative. In general, depending upon intensity, the increase in mean particle size for this black in a filmless preparation is of the

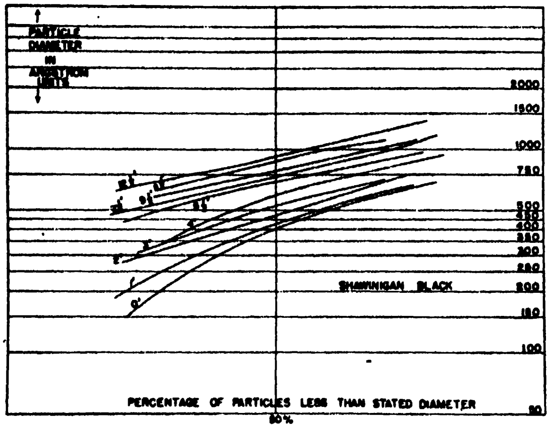


FIG. 2. Particle size frequency distribution curves for each of the exposures in Fig. 1, respectively.

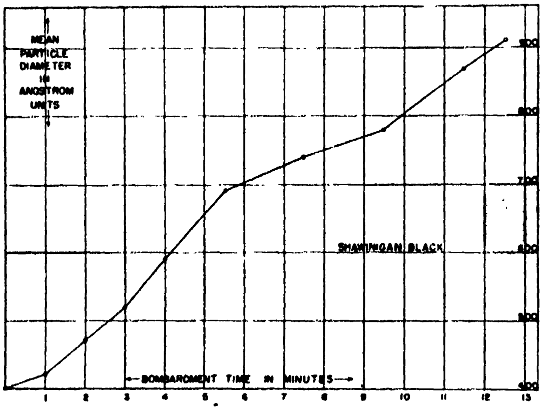


FIG. 3. Mean particle size for each of the exposures in Fig. 1, respectively, plotted against time of bombardment relative to the first exposure Fig. 1A.

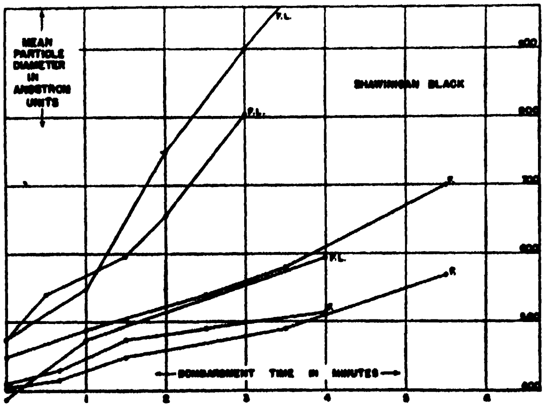


FIG. 4. Other typical increases of mean particle size with bombardment time for filmless and film preparations of Shawinigan Acetylene Black. F., film preparation; F.L., filmless preparation.

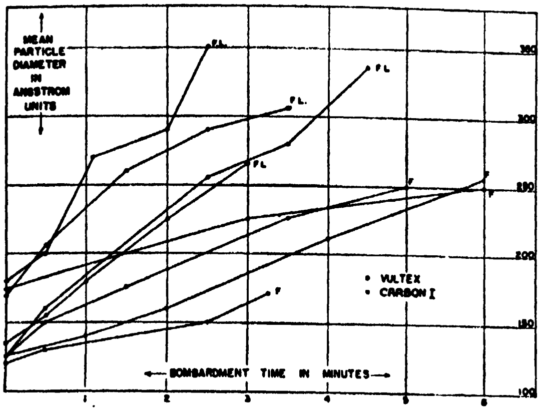


FIG. 5. Typical increases of mean particle size with bombardment time for filmless and film preparations of Voltex and Carbon I. F., film preparation; F.L., filmless preparation.

order of at least 70 percent after 5 minutes bombardment. This is easily a significant value for the type of counting procedure involved. At fairly high, although not focused intensities, the particle size usually will have changed significantly (i.e., say more than 5 percent) in 30 or sometimes in as little as 15 seconds.

In studying a bombardment time series, the same field is photographed after discrete intervals and particle size frequency distributions are obtained from each exposure for comparison. As an example, the frequency distributions, plotted upon logarithmic probability paper[4] (the distributions being negatively skewed), are given in Fig. 2 for each of the exposures in Fig. 1, respectively. In the exposures of this series, too few particles are visible to give good straight lines. The mean particle diameters thus obtained are plotted against time to give the graph in Fig. 3. Other typical increases for acetylene black are shown in Fig. 4. The mean values in the first exposures are within ±10 percent of their arithmetic mean which is fair agreement considering the fact that but one field of the specimen is under study in each case and the number of particles counted is necessarily low. These curves are labeled as being from either filmless or film preparation.* For any particular field, if the intensity is kept constant throughout a run, the relationship between mean particle diameter and bombardment time follows a straight line.

Similar bombardment experiments have been carried out using specimens of such well-known colloidal carbon blacks as Micronex, Voltex, Statex, and P-33. These are structureless carbons as compared with acetylene black which is characterized particularly by its chain structure. Micronex and Voltex are representative of channel blacks of very small diameter. Statex is close to Shawinigan in mean size but is a combustion furnace black, where Shawinigan is formed by a process of thermal decomposition. P-33 is a large diameter black formed by a thermal furnace process. Table I gives (a)—particle sizes for each black represented, (b)—a brief description of the process by which the black is made, and (c) a summary of average values in angstrom units of the increases which can be expected for film and filmless preparations over a 3-minute bombardment period at normal focusing intensity. This table is merely intended to give the reader some idea of expected variation in particle sizes. Other than this it has little significance since the contamination is dependent, as will be discussed later, on several factors which it is difficult to control. Figure 5 gives several curves for mean particle diameter *versus* bombardment time for Carbon I and Voltex. Again, the mean sizes reproduce well from the *first* exposures of the series. Figure 6 shows a typical sequence of Voltex under the electron beam for a film preparation, and

TABLE I. Average increase in mean particle diameter after three minutes of bombardment at normal focusing intensity. The distance of the material from the wire of the grid is unknown.

| Sample | Process of manufacture | Mean* particle diameter | Film mounting | Filmless mounting |
|---|---|---|---|---|
| | | (*A*) | (*A*) | (*A*) |
| Carbon I | Channel Black | 140 | 55 | 155 |
| Voltex | Channel Black | 155 | 110 | 155 |
| Micronex W-6 (EPC) | Channel Black | 295 | 90 | 200 |
| Statex | Furnace Combustion Black | 400 | 115 | 240 |
| Shawinigan | Thermal Decomposition of Acetylene | 430 | 115 | 345 |
| P-33 | Furnace Thermal Black | 1050 | Negligible except at high intensities | |

* The mean particle diameters published here are calculated as an average of the values obtained in these experiments from the particle measurements made on the *first* exposures of the series. The *longest* diameters are taken when measuring a particle. The effect of distortion has been kept as far as possible at a minimum.

---

[4] J. M. Dallavalle, *Micromeritics* (Pitman Publishing Corporation, New York, 1943), p. 47.

* The term "filmless preparation" is described under reference 3. "Film" preparation referred to here involves blowing the dry carbon from a de Vilbuss atomizer toward a Formvar film over a usual electron microscope specimen holder.

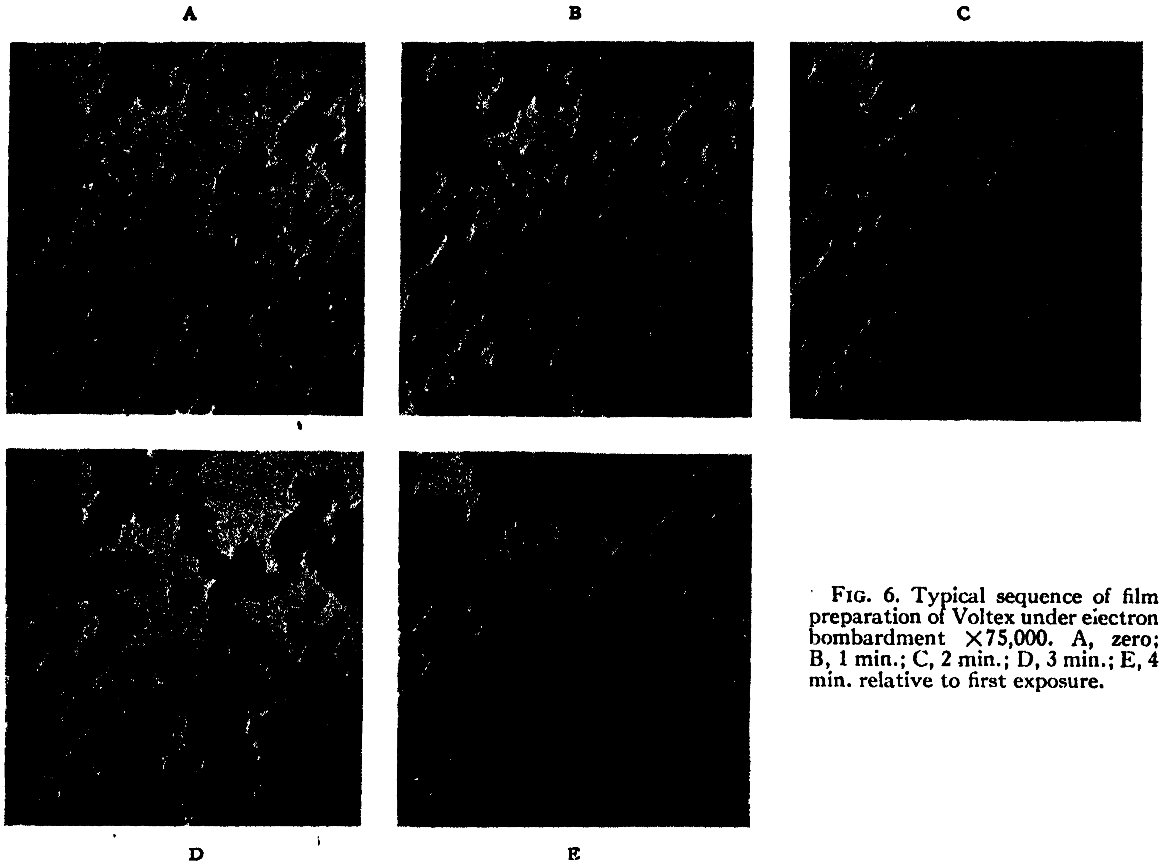


Fig. 6. Typical sequence of film preparation of Voltex under electron bombardment ×75,000. A, zero; B, 1 min.; C, 2 min.; D, 3 min.; E, 4 min. relative to first exposure.

Fig. 7 for a filmless one. Increases of well over 100 percent may be recorded for film specimens of the finest carbons such as Carbon I and Voltex in 3 minutes. From filmless preparations of these two carbons it is difficult to secure particle diameters at all after a few minutes since the clumps seem to coalesce and single particles are no longer visible.

Similar changes are observed with Statex samples and to a marked lesser degree with P-33. However, with very intense bombardment of P-33, large changes are recorded. In general, as would be expected, the percentage increase in mean particle diameter is greatest for blacks of smaller particle size.

Several general observations have been made concerning the relative magnitude of the effect under different conditions of specimen examination. It is more pronounced in a filmless rather than a film mounting. It is zero when the specimen is incorporated into the body of a Formvar film. It is most pronounced at high intensities and may be made negligible if the intensity is lowered sufficiently. With prolonged bombardment of carbon black the effect does not clean up but is permanent, and becomes progressively more pronounced as shown in Fig. 1. The effect is of equal magnitude regardless of whether the sample is on the filament side of the supporting film, or on the reverse side. The carbon "grows" only in that area where the specimen is being bombarded.

That a contaminant surface is deposited, is easily demonstrated by bombardment of a Formvar film which has holes through its surface. The film thickens noticeably and contrast between film and hole space is enhanced by the bombardment, see Fig. 8. Figure 8B was taken after a 3-minute bombardment with all operating conditions and photographic processes, the same as for Fig. 8A. This film thickening will lower resolution and contrast over the specimen but not to a sufficient extent in short times to account for the pronounced increases in mean particle size recorded here. The holes in the Formvar are

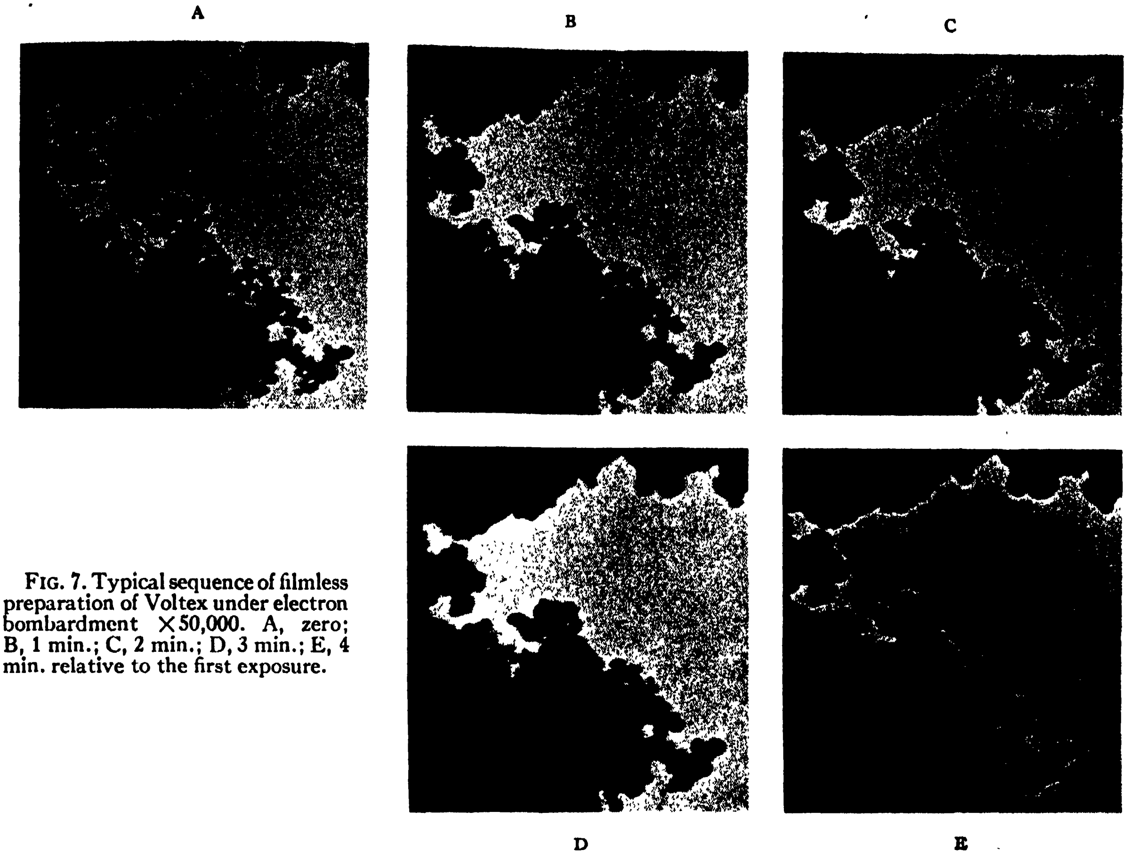


FIG. 7. Typical sequence of filmless preparation of Voltex under electron bombardment ×50,000. A, zero; B, 1 min.; C, 2 min.; D, 3 min.; E, 4 min. relative to the first exposure.

measurably smaller after bombardment; a decrease of about 8 percent occurring in 5 minutes.

It has been pointed out already that a greater contamination effect is observed in filmless than in film preparations. In a filmless one a field is usually chosen which is quite near the metal of the specimen grid. Upon further investigation, it is observed that in either type of specimen preparation, the closer the bombarded area is to the grid wire the greater is the effect of contamination. Figures 9 and 10 are offered to illustrate this. These were taken of the *same* specimen of Shawinigan Black mounted on a filmless holder. Figures 9 A and B were taken of a field in the center of one of the holes, as far from the wire as possible. No change was noted after a bombardment of six minutes. Figures 10 A and B were taken of a field immediately adjacent to the wire. There is an 81 percent change in mean particle size between these two figures after one minute bombardment. The mean particle size changes from 470 to 850 A.

An attempt is being made to relate percentage change in mean particle size with distance from the screen. From preliminary experiments it appears that the very pronounced contamination falls off rapidly and only occurs in a region distant about 8$\mu$ from the wire. This distance will vary with different specimens and electron intensities. In regions further than this from the screen the contaminant builds up more slowly. In order to keep the effect at a minimum, fields for examination should be chosen as far as possible from the metal grid.

The observations cited support an explanation based upon specimen contamination occasioned by the deposition of a hydrocarbon material under electron bombardment after the manner described in reference 2. The mechanism of this hydrocarbon formation in the electron microscope is related to the well-known phenomenon of polymerization and condensation of organic vapors and gases, and certain other gaseous products, under electron or other charged-particle

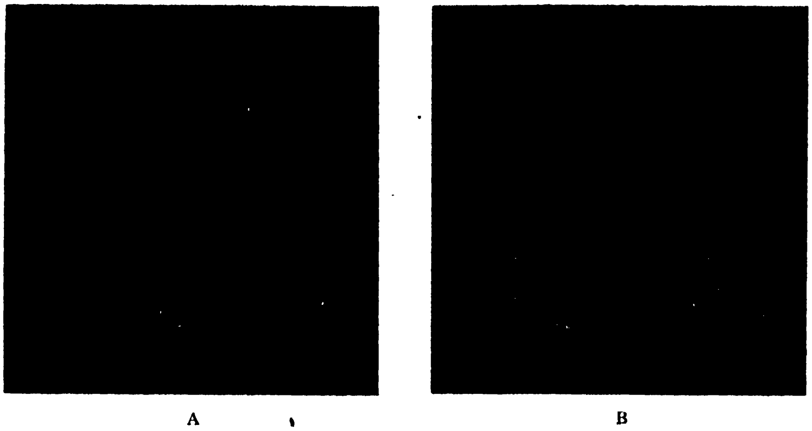


FIG. 8. A, Formvar film before bombardment. B, Same field after three-minutes bombardment ×25,000.

bombardment.[5,6] Theoretically these contaminant surfaces may originate either from organic materials in the metal body of the microscope itself, from grease used to a limited extent on vacuum seals, or from impurities present in the specimen holder and supporting grid. It is con-

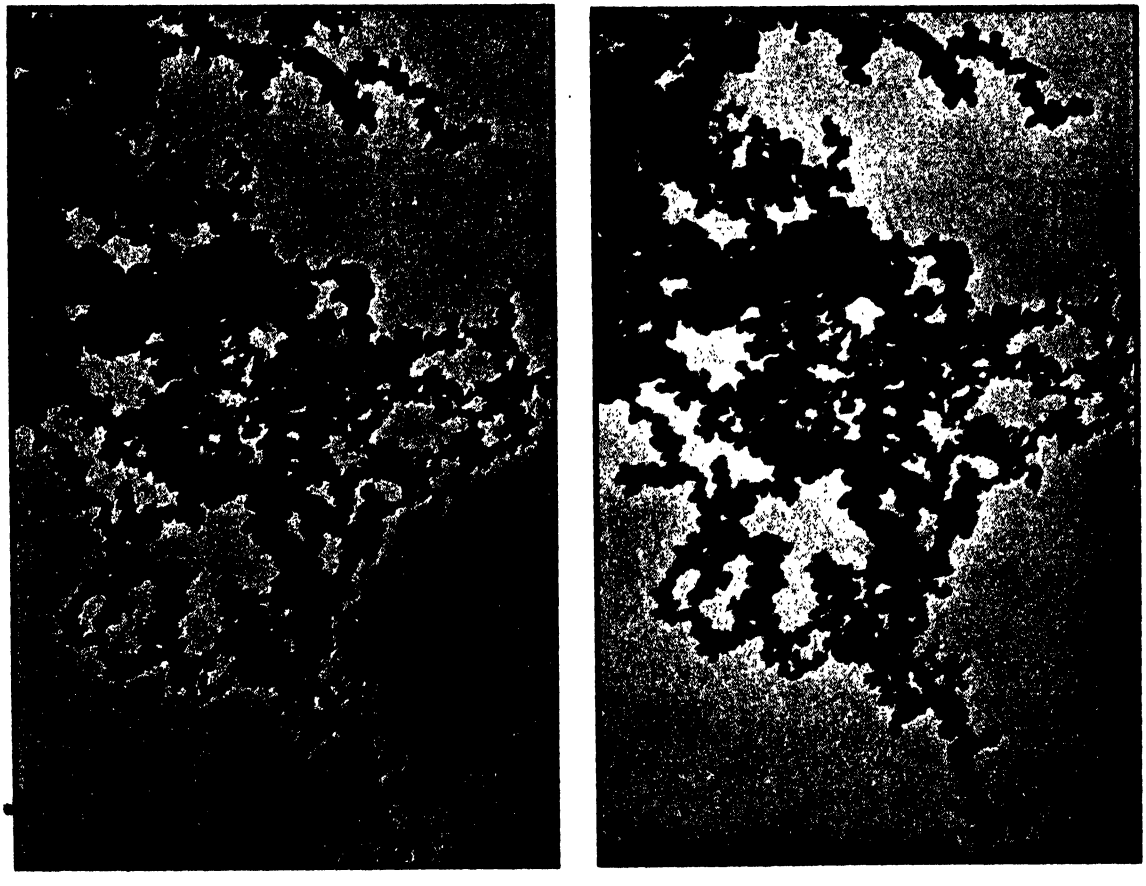


FIG. 9. Shawinigan Carbon Black, filmless preparation, micrograph taken of a field which was as far from the metal wire of the grid as possible. A, no bombardment. B, after 6 minutes of bombardment.

---

[5] George Glockler and Alvin E. Walz, Trans. Electrochem. Soc. 88, 151–160 (1945).

[6] J. C. McLennan, M. W. Perrin, and H. J. C. Ireton, Proc. Roy. Soc., A125, 246 (1929).

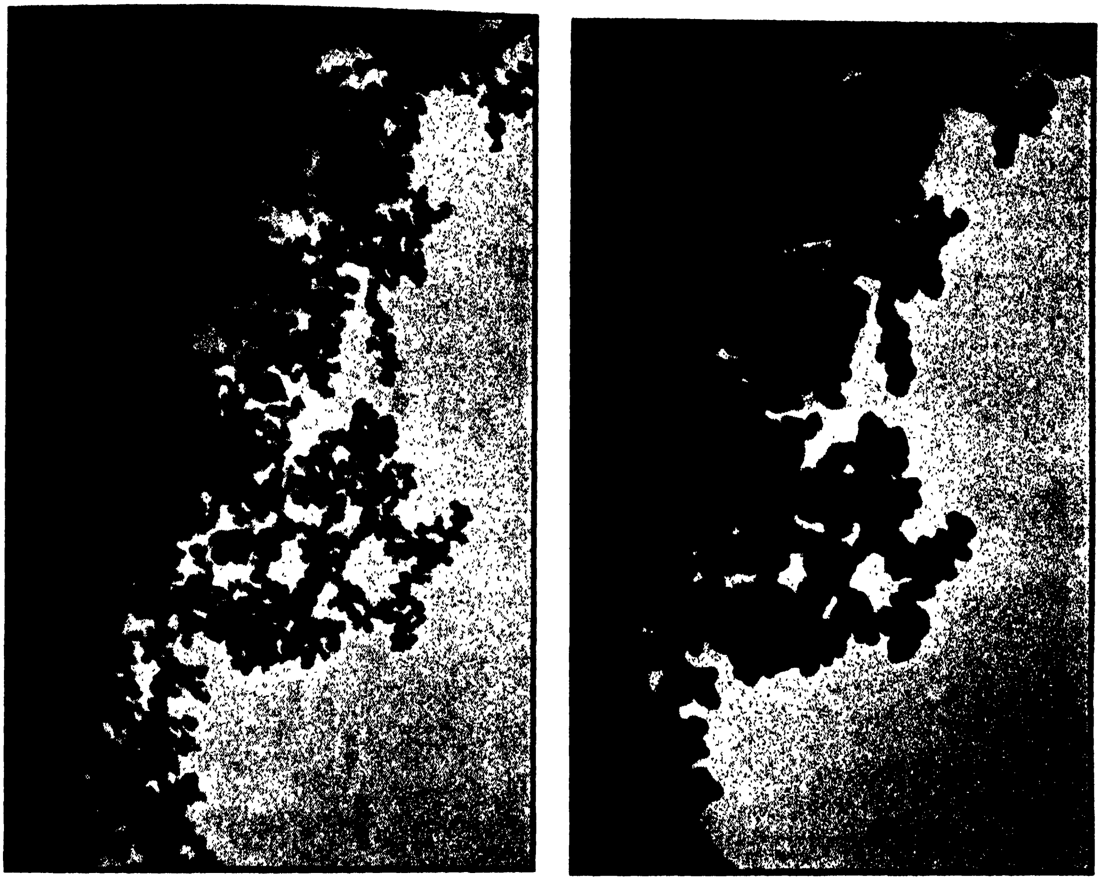


FIG. 10. Same specimen as Fig. 9, micrograph taken under same conditions as Fig. 9 of a field immediately adjacent to the grid. A, no bombardment. B, after 1-minute bombardment.

ceivable also that organic substances present in the object could be distributed as a vapor under the beam and be redeposited upon the particles; but it is known that the percentage of volatile material in these blacks bears no relation to the magnitude of the change.

If we assume that the explanation for the contamination lies in the above paragraph, a reason for the more pronounced effect near the grid can be advanced. The formation of the polymer can be explained from Lind and Glockler's[7, 8] experiments upon alpha-ray bombardment and consequent polymerization of acetylene and other hydrocarbons. They suggest that when the molecules in gaseous or vapor state are bombarded by charged particles they are ionized and that they then attract neutral molecules to themselves to form a cluster-ion. When this finally is neutralized a large molecule is formed.[9] This neutralization and resulting deposition of the contaminant will occur most easily upon electron microscope specimens at or near the *conducting* grid.

In order to test this explanation a conductive film of chromium metal was deposited over a Formvar film preparation of Voltex Carbon, by use of shadow casting methods. Under these conditions, marked contamination was observed over the specimen regardless of position relative to the grid. Figures 11 A and B are reproduced of a field

---

[7] S. C. Lind, and G. Glockler, *The Electrochemistry of Gases and Other Dielectrics* (John Wiley and Sons, Inc., New York, 1939).

[8] G. Glockler and S. C. Lind, J. Am. Chem. Soc. 53, 3355 (1931).

[9] J. H. L. Watson and K. Kaufmann, J. App. Phys. 17, 996 (1946).

approximately in the center of a grid hole. The normal degeneration of contrast over the specimen and increase in mean particle size is observed in Fig. 11 B.

Carbon black, and especially Shawinigan Acetylene Black, is particularly conductive. Heavy contaminant deposits would be expected

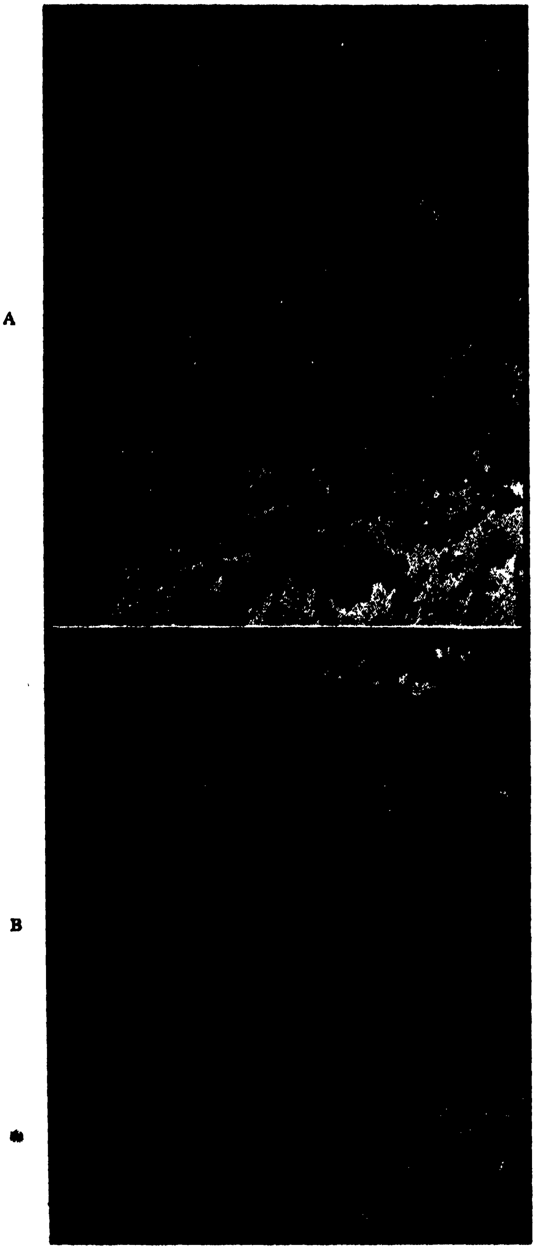


FIG. 11. A chromium shadow cast specimen of Voltex. A, no bombardment. B, after a 4-minute bombardment. The field was chosen far from the grid wire.

therefore to form upon it. Whereas it is always possible with carbon black to demonstrate a rapid increase in size under bombardment, it has not been found possible in this laboratory to do so to the same extent with other materials.

The loss of contrast over the specimen and the lack of detail are explainable on a basis of contamination of both the film and the specimen itself. Likewise masking of the small particles; growth of all the particles; and loss of original shape characteristics could be caused by contamination.

That as large an effect is not noted with other substances as with carbon black may be significant also as an indication that the surface of the black has a certain activity for attracting and holding the contaminant permanently so that it will not clear up under continued bombardment. This would conform with the known activity of carbon surfaces in other regards[10] such as their adsorbent, catalytic, and unsaturated electron properties.

Recognition of the existence of the effect points out at least one way in which values from particle size frequency distribution curves can be secured with greater accuracy and reproducibility. Since increases in particle sizes similar to these but of less magnitude have been found to occur to some extent with other materials under bombardment, precautions for minimizing the effect should be followed in all particle size determinations made with electron microscopes.

The change takes place so rapidly with finer carbons that in any micrograph where there is a possibility that it has occurred the *original shape* of the particles is masked almost certainly. Therefore it may distort not only particle size determinations but it may lead to wrong conclusions concerning particle shape and through these to incorrect calculations of specific surfaces. In Shawinigan Acetylene Black for example, the fine, platelet particles appear as round or oval units after a few minutes. In experiments to date the finest carbons such as Carbon I and Voltex appear to be "feathery" in the original uncontaminated condition but rapidly become small, distinct spheres under bombardment. The feathery appearance may be caused wholly by

---

[10] U. Hofmann and W. Höper, Naturwiss. **32,** 225–226 (1944).

lack of resolution but the possibility exist that some of these extremely fine carbon particles possess crystalline properties and straight-edged shapes in their original condition, much in the manner of Shawinigan Acetylene Black. The shape characteristics however would be within, or near, the limit of resolution, and would not be seen except insofar as they lend the feathery appearance to the image.

The effect may be made zero by incorporating the black samples into the supporting film. However, the sample manipulation required in effecting this type of mounting will tend to change the appearance of the carbon from what it was originally, causing structure breakdown and making interpretation of results more difficult and less reliable. In this laboratory it is desirable to study the black as much in its normal state as possible and therefore dry mounts are essential. It is in these that the change under bombardment is considerable. Therefore it becomes a necessity that conditions in the microscope be kept such that a minimum of change is possible. To do this the following rules of procedure are followed: (1)—the intensity is kept at a minimum, (2)—micrographs are taken as rapidly as is practically possible, (3)—focusing is often carried out using a part of the field which is not to be photographed, and (4)—fields are photographed which are as far as possible from the metal of the specimen grid. If a black is bombarded at fairly high intensity, (focusing intensity) for any length of time longer than 15 seconds a correction downward in particle size should be applied. It is impossible to give a value for this correction here as it will depend upon the time taken for focusing and the intensity used by the individual microscopists.

Many of the observations and much of the information concerning contamination is contradictory and incomplete. Since it is becoming an increasingly important problem in electron microscopy an exchange of information on the subject in publication is needed in order to clarify many aspects of it.

# On Cavity Formation in Water*

E. Newton Harvey, Wm. D. McElroy, and A. H. Whiteley
*Department of Biology, Princeton University*

(Received October 28, 1946)

**Phenomena involving the tensile strength of water have been studied by a kinetic method—high speed motion photography of the rapid movement of a blunt glass rod (5 mm diameter) in a narrow (16 mm inside diameter) glass tube of water. Special precautions have been taken to remove all hydrophobic patches and small gas masses (gas nuclei) but to retain the dissolved gas (air at one atmosphere) in the water. If the rod surface contained gas nuclei, or was hydrophobic and free of gas nuclei, cavitation occurred at the rear end when the velocity was less than 3 meters/sec., but if completely hydrophilic and free of gas nuclei, the velocity could be 37 meters/sec. or 83 miles/hour without cavitation. Addition of a detergent (diactyl sodium succinate) to the water did not prevent cavitation at a low velocity with the hydrophobic rod free of gas nuclei. Movement of a rod in pure corn syrup (viscosity 20.1 poises), free of gas nuclei, left a large cylindrical cavity that collapsed in a matter of hundredths of a second. It is not possible to calculate the tensions developed in these experiments, but it is pointed out that the velocities attained without cavitation are far higher than previously observed for movement of bodies in an aqueous medium, a result believed to be owing to the absence of all gas phases and hydrophobic surfaces.**

## INTRODUCTION

During a study of bubble formation in the blood and tissues of animals—a condition leading to such symptoms of decompression sickness as the "bends"—the conclusion was reached that a decrease in hydrostatic pressure, i.e., a tension in the tissues, was one of the primary factors involved. Favoring this view is the observation that straining movements and muscular exercise at simulated high altitudes greatly increase the incidence of bends in men or of bubbles in animals.[1] This work led to investigation of the tensile strength of pure water containing dissolved air, and it is the purpose of this paper to describe the method and certain observations in connection with the experiments. Calculation of the actual tensions involved has not been attempted, but all relevant data are given.

There is a general belief that water under tension can be easily torn apart; that it is only necessary to reduce the pressure to that of water vapor in order to obtain profuse bubbling and boiling. Actually, the boiling which occurs at the vapor pressure is because of the presence of minute gas pockets adhering to dust particles in the water or on the walls of the container. The gas phase serves as a gas nucleus to start the cavitation. It is a simple matter to show that if gas nuclei are removed, water in non-capillary spaces can be subjected to high vacua or can be heated to over 206°C without boiling,[2] although evaporation from the surface is rapid (Harvey *et al.*[3]).

Many experiments, recently reviewed by Vincent[4] and discussed by Dean[5] indicate that a water column in a well-cleaned container can be subjected to a tension measured in atmospheres without breaking. The maximum values for this tension have been obtained by means of the thermal contraction method. In this method a glass tube, *completely* filled with the liquid at a higher temperature, is slowly cooled. The differential contraction of liquid and container then exerts the tension that finally pulls liquid from the walls with a sudden click. Dixon[6] obtained 150 atmospheres and Vincent[4] 157 atmospheres, using this method.

Both values are far short of the 10,000 atmos-

---

[*] The work described in this paper was done under a contract recommended by the Committee on Medical Research, between the Office of Scientific Research and Development and Princeton University.

[1] For a general discussion of decompression sickness and bubble formation, see E. N. Harvey, Bull. N. Y. Acad. Med. 21, 505 (1944).

[2] The vapor pressure of water at 206°C is 17.3 atmospheres and at 270°C, the highest temperature to which Kenrick, Gilbert, and Wismer, J. Phys. Chem. 28, 1297 (1924) could superheat water, is 54 atmospheres.

[3] E. N. Harvey, A. H. Whiteley, W. D. McElroy, D. C. Pease, and D. K. Barnes, J. Cell. Comp. Physiol. 24, 23 (1944).

[4] R. S. Vincent, Proc. Phys. Soc. London 53, 126 (1941) and 55, 41 (1943).

[5] R. B. Dean, J. App. Phys. 15, 446 (1944).

[6] H. H. Dixon, Proc. Roy. Dublin Soc. 14 (N.S.), 229 (1914).

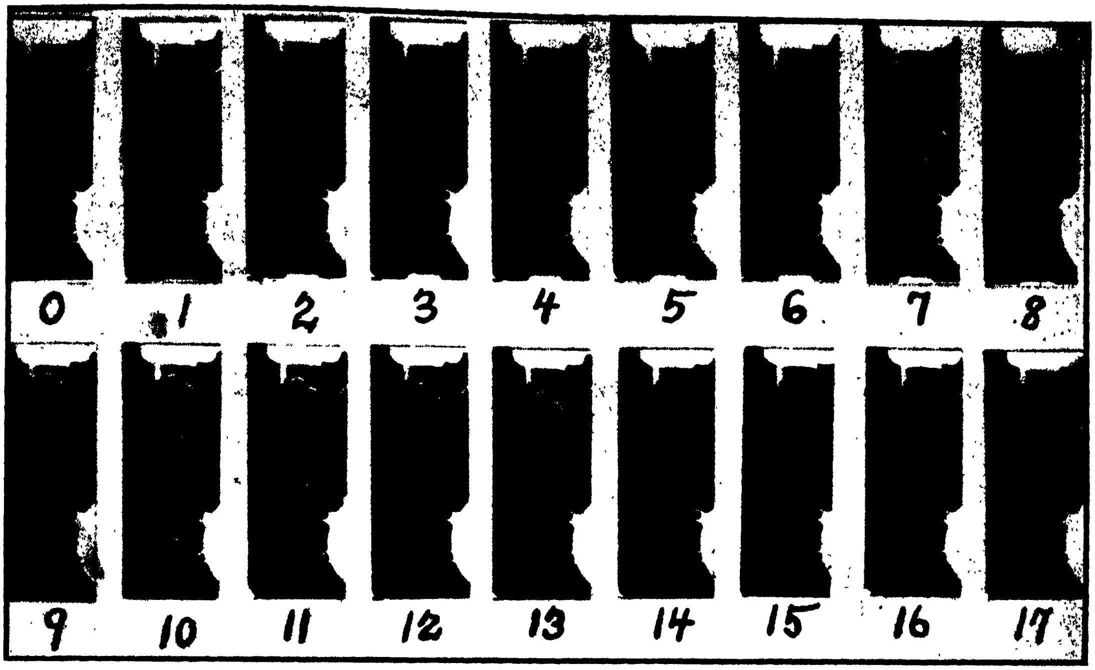


FIG. 1. Moving picture of cavity formation due to a pressure pulse (with negative pressure component) which arises when the bottom of a glass tube containing highly evacuated water (vapor pressure = 19 mm Hg) is struck a blow with a wooden block. The block appears in frame 1 and disappears in frame 10. Note that the cavities, which have enlarged from gas nuclei, first become of maximum size near the bottom of the tube and last longest near the top of the tube. Tube diameter 19.3 mm, 140 frames per second. Exp. 6-9-43.

pheres for the mutual attraction between the water molecules, or the "intrinsic pressure" within the liquid, possibly because breaking is not simultaneous across a unit cross-sectional area. On the other hand, the variation in experimental values from 5 to 157 atmospheres that have been obtained by different workers, may be a reflection of insufficient removal of small gas masses.* In every experimental investigation of the tensile strength of water the container is as important as the water. Not only must there be no gas nuclei on the walls but also no hydrophobic surfaces must be present, since water does not adhere to these surfaces and readily separates from them. Both these conditions indicate that a relatively small amount of water and a minimum surface area of easily cleaned hydrophilic material is essential for experimental work. Glass is unquestionably ideal for the surface since the water-glass contact angle is zero.

In addition to the previously mentioned thermal contraction method of exerting tension on water, kinetic methods are also available. In one of these the bottom of a tall glass tube filled with water, which has been exhausted to the vapor pressure, is struck a sudden blow. The blow moves the glass wall of the tube sufficiently to start a pressure pulse with a negative component. The cavitation and bubble formation under those conditions are well known. In Fig. 1 is reproduced a series of prints from a high speed moving picture which show the resulting cavitation when the bottom of a tube of evacuated water (vapor pressure = 19 mm Hg) is hit with a wooden block. The cavities have started from gas nuclei.

Another method of developing tension is to move a long tube of water rapidly in the direction of its open end and suddenly bring it against a rigid barrier or stop. The momentum of the moving water tends to pull it apart.

Although the above two procedures have readily caused cavitation and bubble formation if gas nuclei were present, after the removal of

---

* Additional references on the tensile strength of water which have recently come to the attention of the author are: A. L. King, Proc. Nat. Acad. Sci. 30, 155 (1944) and H. N. V. Temperley and L. G. Chambers, Proc. Phys. Soc. London 58, 420, 436 (1946).

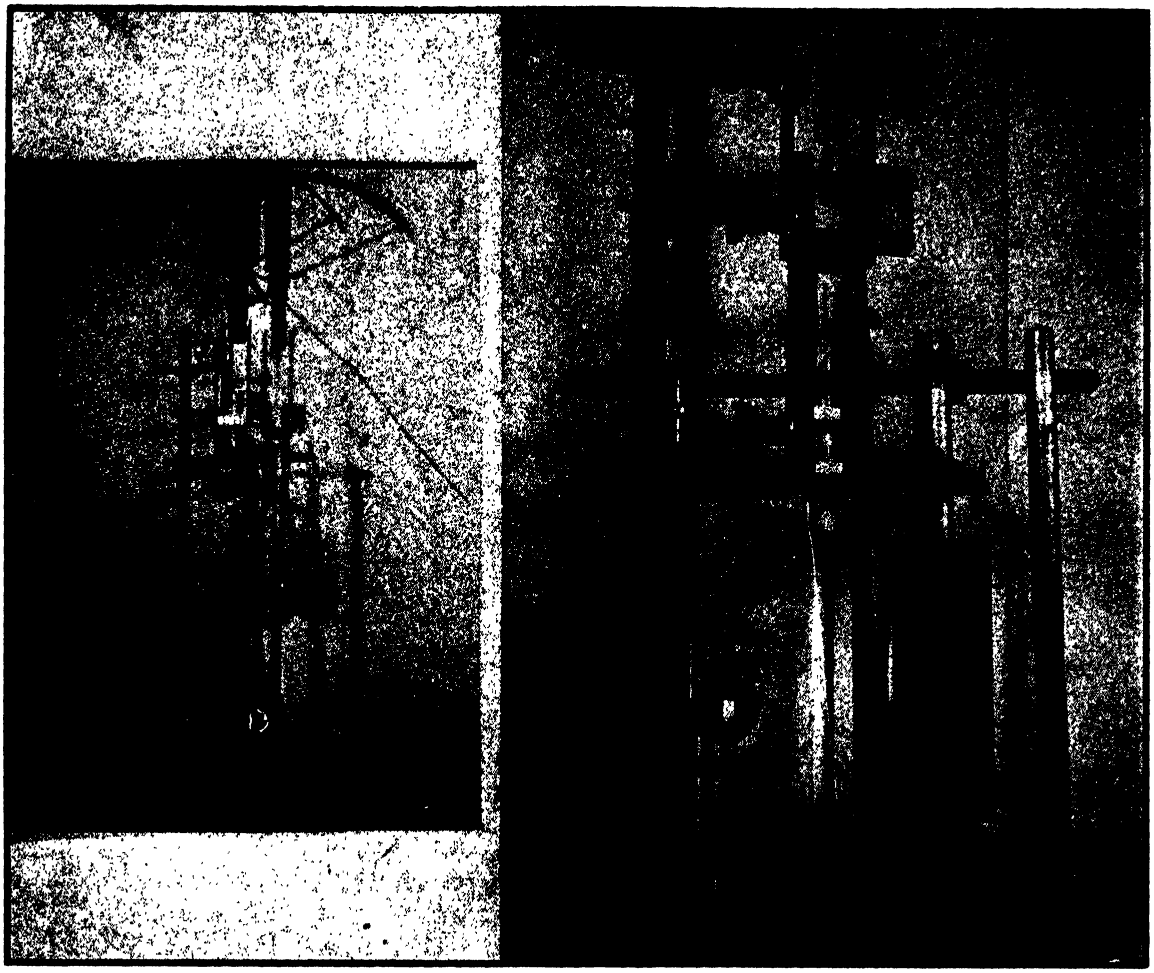


FIG. 2. A far (left) and near (right) view of the apparatus for rapidly moving a glass rod vertically in water. The steel bow appears above and the sodium lamp (with a circular mirror) on right. This lamp flashes 120 times a second and is used for timing the frames. In the near view, the Cellophane covering of the tube to protect against dust and a large mirror to allow a view of the cavitation at right angles to the camera is plainly seen.

the gas nuclei the blow to the tube or the velocity of the movement necessary to cause cavitation were so great as frequently to break the tube. Likewise we have found that cavitation of water, caused by the passage of high frequency sound waves, ceased when gas nuclei were previously removed, although a fairly high intensity of sound was employed (Harvey *et al.*[8]).

Flow methods of reducing pressure, which depend on the principle of Bernouilli (a Venturi or a Reynolds tube, Fig. 6) or the flow past a fixed body in the stream, are not feasible because of the large amount of water necessary and the problem of removing gas nuclei from such a volume. However, by moving a surface rapidly through stationary water similar regions of low pressure develop. The rear end of a moving object, such as a rod, exerts a continuous pull on the surrounding medium, which attempts to follow the rod and close the potential space. This method has been adopted in the experiments described below.

### TECHNIQUE

In order to study the tearing of water from a glass surface moving with high speed, conditions most favorable for the development of tension were selected, namely, a cylindrical glass rod cut squarely across at one end and moved rapidly through a small volume of water in a narrow glass container. The walls of the container were near enough to those of the rod to give the additional suctional effect of a leaky piston. Rod and container cannot be too close, as the contact of one surface with another supplies the conditions for immediate cavity and bubble formation (Harvey *et al.*[7]).

To make certain that no hydrophobic spots remained on the glass surface, cleaning was accomplished by washing with soap and water, followed by hot trisodium phosphate and then hot concentrated sulphuric acid-bichromate mixture. Dust particles were removed by centrifuging (800×g for 30 minutes) the water used in the experiment. Our experiments differ from those of previous workers in that attempts were made to remove completely all gas nuclei, but to leave the dissolved air (at a tension of one atmosphere) in the water.

The method for removing gas masses was that of high hydrostatic pressure treatment (Harvey *et al.*[3]) which forces gas nuclei into solution. The glass tube, 16.0 mm inside diameter and about 41 cm long, with its contained water (distilled) and glass rod in position, with wire attached for pulling the rod, was covered with dialyzing Cellophane tubing tied both to the wire and to the glass tube end to prevent the entrance of dust, as shown in Fig. 2. This assembly was then placed in water in a long steel pressure chamber and subjected to 16,000 lbs./in.$^2$ hydrostatic pressure for from 15 to 120 minutes.

After the cleaning and pressure treatment, the tube, rod, and water were tested by evacuation to the vapor pressure of water to see if bubbles would arise. In no case was bubbling observed. The assembly was then mounted in a frame and the rod connected by its wire either to a coiled wire spring (in earlier experiments) or a bow, fabricated from a steel auto spring leaf, for rapid movement of the rod.

The lower part of the tube and rod were arranged in position before an Eastman 16 mm high speed moving picture camera, the maximum speed of which is 3000 frames per second. Behind the tubes there was a ground glass, illuminated by two No. 2 photo-flood lamps, run on 130 volts in a reflector. This intensity of transmitted light was found to give proper exposure for 2000 frames per second with $f=3.5$ and a camera distance of three feet. In some experiments a mirror was so arranged that a view of the moving rod at right angles, in addition to the direct view, was recorded by the camera. The bow or spring was stretched and locked in position with a trigger release that was activated when about 35 feet of a 50-foot roll of Eastman super XX panchromatic film had passed behind the camera lens. The film was developed in D11 for 5.5 minutes. Just before the motion picture of the moving rod was taken, the Cellophane cover, which protects from dust, was loosened around the edge of the glass tube but left attached to the rod wire, thereby moving upward with the rod. No dust could enter the water until after the rod had moved out of the camera field of view.

Two sizes of Pyrex glass rod have been used, both about 29 cm long, one 5 mm in diameter, and one 10 mm in diameter. The end of the 5-mm rod was fire polished while the end of the 10-mm rod was ground and polished with rouge to make a right angle. The larger rod was introduced to increase the suction and tension on the water. The wall clearance between the 10-mm rod and tube (16-mm internal diameter) is only 3 mm, and the ratio of cross-sectional area of rod (78.5 mm$^2$) to cross-sectional area of tube (201 mm$^2$) is 0.39. It is not feasible to increase this ratio further because scraping of glass on glass causes bubble formation in gas supersaturated solutions or in a liquid under reduced hydrostatic pressure. Such a scraping of glass rod on glass wall has never occurred in our experiments but would undoubtedly result in cavitation.

## EXPERIMENTS

To demonstrate the tearing of water from a hydrophobic surface, dry clean glass rods 5 mm in diameter were covered with a thin film of paraffin or a thin layer of General Electric repellent varnish[8] (two experiments). The special cleaning procedure could not be applied to these surfaces. They were placed in the tube containing

---

[7] E. N. Harvey, K. W. Cooper, and A. H. Whiteley, "Bubble formation from contact of surfaces," J. Am. Chem. Soc. 68, 2119 (1946).

[8] We express our thanks to Dr. F. J. Norton, of the General Electric Co. for a sample of this varnish.

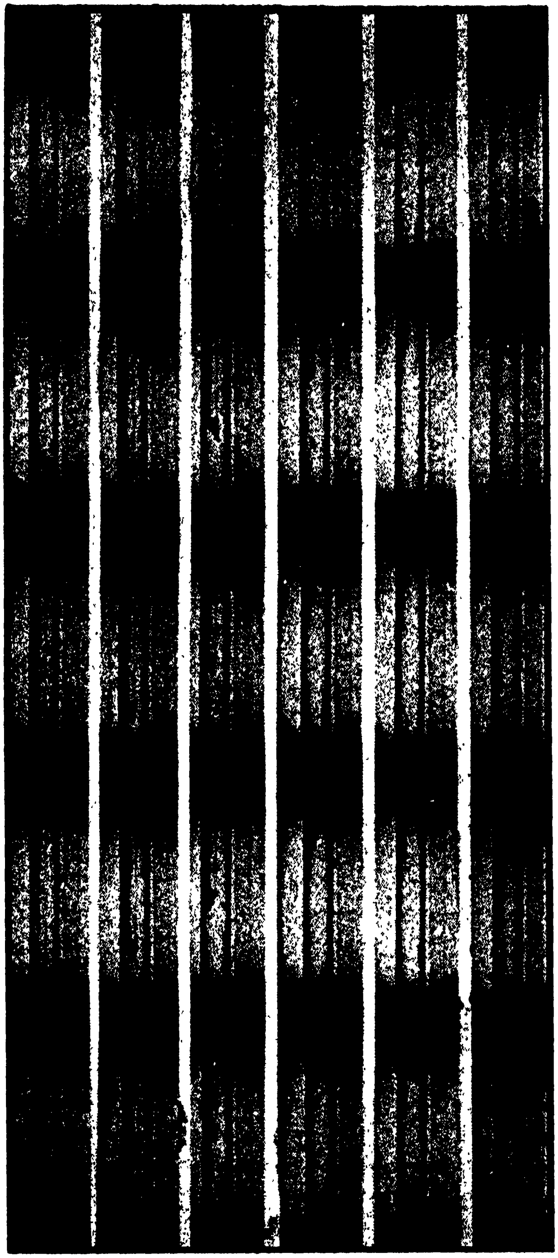


FIG. 3. Moving picture (2280 frames/sec.) of a paraffin tipped rod (5-mm diameter) moved through gas nucleus-free water at a final maximum speed of about 16 meters a second. Note the abundant cavitation, characteristic of hydrophobic surfaces, which appears in the first frame when the velocity was less than 3 meters a second. Cavitation disappears in frame 17, to reappear and again disappears in frame 22, with subsequent periodicity. Scale in centimeters. Exp. 9-7-44.

water and pressurized at 16,000 lbs./in.$^2$ for from 30 to 90 minutes. When tested at the vapor pressure of water (20 mm Hg) no bubbles arose from the glass surfaces or from the water. However, motion pictures of the movement of the hydrophobic rods through water showed cavitation on the first frame when the velocity was less than 3 meters/sec., and a misty trail of fine cavities in succeeding frames, as shown in Fig. 3. The water in the tube after the rod was drawn out was observed to be full of fine bubbles. On the other hand, completely hydrophilic rods, 5 mm in diameter, have been drawn through the water with a velocity of 37 meters/sec. without a sign of cavitation. A plot of rod velocity *vs.* time is shown in Fig. 4 and will illustrate the rapid acceleration attained with our type of bow.

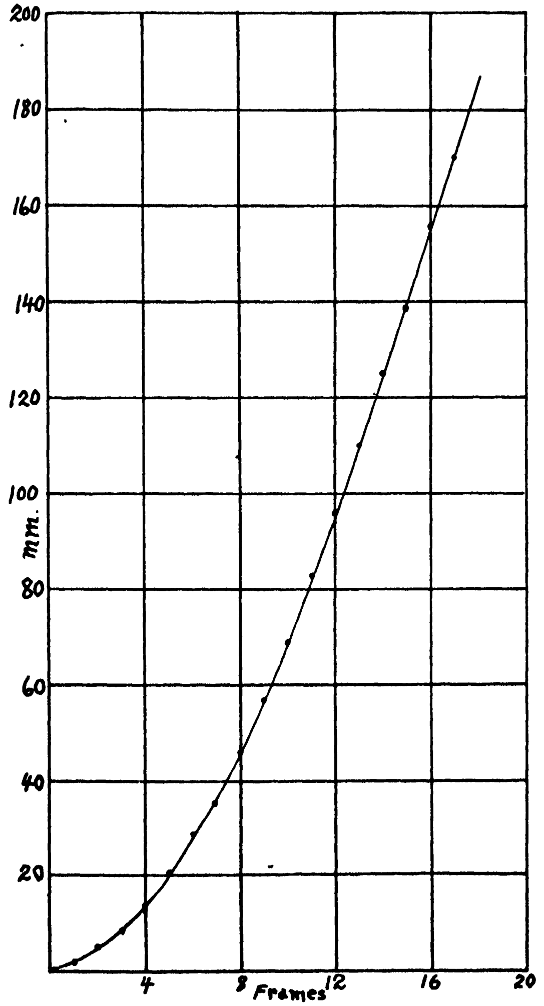


FIG. 4. A plot of the acceleration of a 5-mm glass rod in gas nucleus-free water. Exp. 12-14-44. 2400 frames/sec. The final velocity was 36.7 meters/sec., yet no cavitation appeared at this speed.

The question arises whether the presence of a detergent in the water, that might cover hydrophobic surfaces, leaving an outwardly directed hydrophilic layer of molecules, might not prevent this ready cavitation. Accordingly, one of the water repellent varnish covered rods (5-mm diameter) was moved rapidly in water containing 0.5 percent aerosol OT (dioctyl sodium succinate). Cavitation appeared practically from the start of movement. It is very possible, however, that movement of the rod tears detergent molecules away from the surface, thereby exposing hydrophobic spots.

The bubble formation on the paraffined rod is very similar to that which is observed if the rod and tube have not been subjected to the cleaning and pressure treatment, i.e., if gas nuclei have not been removed. Again the first frame shows a fine mist of cavitation at the end of the rod, when the velocity may be only 3 meters/sec.

In some of the early experiments, where cleaning of glass surfaces involved soaking in cold concentrated sulphuric-bichromate mixtures only, and the tube was not protected from dust particles, cavitation occurred when the 5-mm rod had reached a high velocity. In one case, shown in Fig. 5, two small cavities appeared on the end of the rod at the 9th frame when the velocity was 18 meters per second. These enlarged until the 12th frame, when they disappeared, but reappeared on the 14th, 15th, and 16th frames. This periodic formation and disappearance of cavities is frequently observed during cavitation, and is particularly well seen in a Venturi nozzle. The sound during such Venturi cavitation is caused by periodic formation and collapse of cavities which were photographed by Hunsaker,[9] using the Edgerton high speed technique. In Fig. 6 are reproduced prints from a 2000 frames per second film of cavitation in a small glass Reynolds tube (a glass tube with an abruptly constricted region), taken with our Eastman high speed camera. The burst of cavitation is periodic although the rate of picture taking is not sufficient to follow the complete cycle.

As in the Reynolds tube, the periodic nature of cavitation below the rod is because of the fact that cavity formation relieves the tension in the water, with consequent collapse of the cavity; then further tension develops with cavity formation, followed again by collapse. The process is

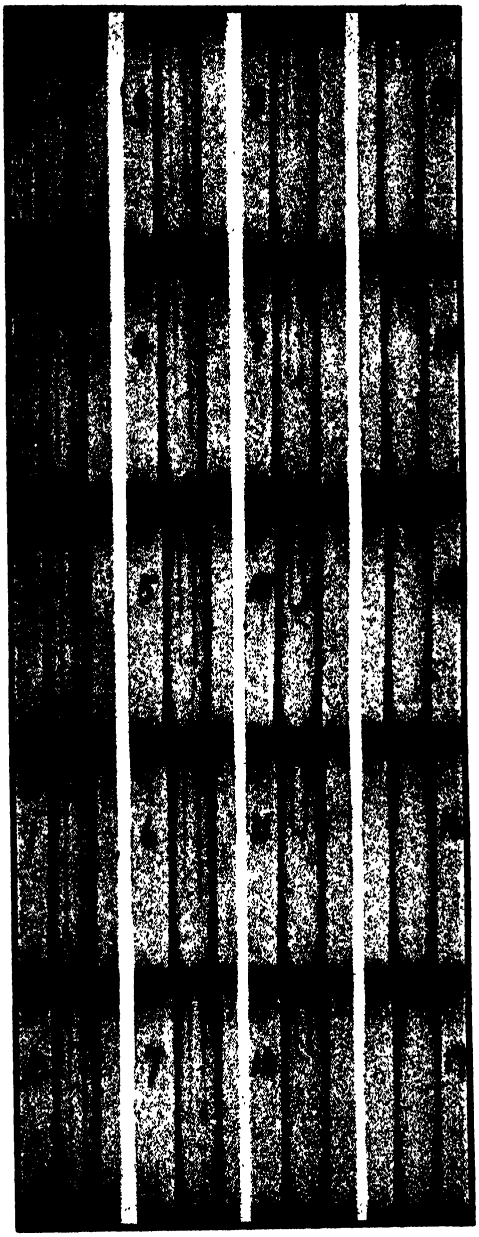


FIG. 5. Moving picture (2280 frames/sec.) of a 5-mm diameter glass rod moved through water in which cavitation first appeared (frame 9) when the velocity was 18 meters/sec. Note disappearance of cavitation in frames 12 and 13, with reappearance in frame 14. External diameter of tube 19.5 mm. Exp. 7-9-44.

---

[9] J. C. Hunsaker, Mech. Eng. 57, 211 (1935).

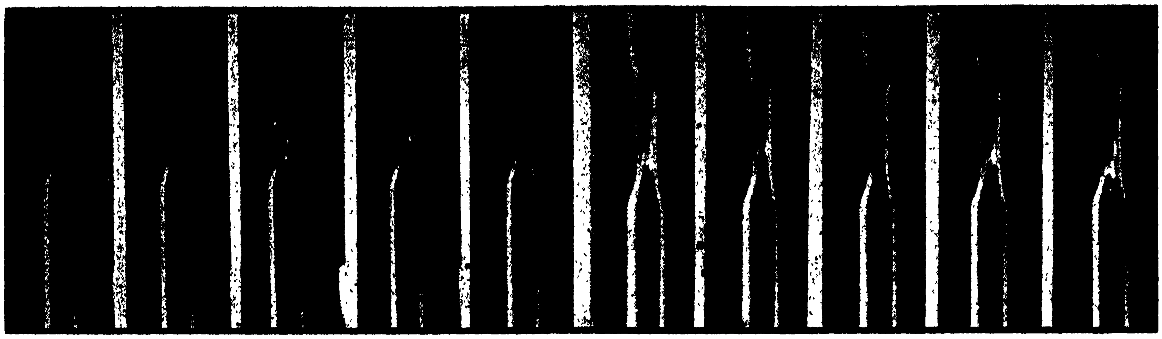


FIG. 6. Negative prints from a moving picture (1680 frames/sec.) showing cavitation in a Reynolds tube (8.3-mm large diameter) from flow (top to bottom) of water through a constriction. Note the periodic cavitation (white in appearance) which always begins at a particular spot on the narrow region of the tube.

repeated as long as the conditions for development of tension persist.

Whenever cavities appear under a rod, they are greatly disturbed by the turbulence and frequently look like a fine mist. If hydrophobic dust particles with air films have fallen into the pressurized water (or are purposely placed there) isolated bubbles may form at a distance from the rod, as shown in Fig. 7. The bubbles may appear almost immediately or after the rod has reached a certain speed. Sometimes they are attached to the wall and sometimes free in the tube. If near the rod, they may enter vortices at the end as the rod passes in the water, and a trail of misty cavitation will follow. Isolated bubbles are also observed to pulsate, appearing for several frames and then disappearing again, reflecting the periodic pressure changes in the tube.

In order to study the effect of viscosity, two 5-mm rods have been drawn from concentrated sugar solutions (Staley's crystal white syrup) at a temperature of 23°C, where the syrup viscosity is 20.1 poises, as compared with a viscosity of 0.00936 poise for water at 23°C. Density of the syrup is 1.383 at 20°C. One rod was of glass and one of aluminum. Since aluminum tends to dissolve in water, this rod might contain films of hydrogen which could not be removed by prepressurization. However, when tested at the vapor pressure of water, the aluminum rod was found to contain no gas nuclei that grew to bubbles. The moving picture film, reproduced in Fig. 8A showed that at the very start of movement (1st frame) several minute cavities appeared at the end and side of the rod and in succeeding frames grew into a long cylindrical cavity the width of which remained approximately that of the rod. Finger-like processes reflecting the initial cavities were present at the lower end where rod movement starts. The cylindrical cavity remained static during the movement of the rod out of the syrup and then started to collapse (Fig. 8B). The collapse occurred along the long axis of the tube from top down, not from side to side, and proceeded at a uniform speed of 940 cm/sec. It was as if the rod had left an impression of itself in the syrup. When the collapse finally occurred, several smaller cavities appeared and oscillated several times before subsiding. The glass rod drawn from the syrup behaved in a similar way, but lacked the several fingers at the tip of the cavity, only one cavity tip appearing.

However, in two experiments, glass rods 5 mm in diameter, moved in 75 percent syrup at temperatures of 25°C and 27°C, failed to cause the development of a cavity, although their velocity reached 18 meters/sec. The viscosity of 75 percent syrup at 25°C is 0.56 poise. There was also no cavitation in citrated cat blood plasma at 18 meters/sec. The viscosity of plasma is only slightly greater than water.

Since the 5-mm rods failed to cavitate in clean pressurized water at the highest velocity (37 meters/sec.) attainable with the steel bow, 10 mm diameter rods were substituted. Both because of the greater cross-sectional surface area at the end of the rod and the greater suction effect in the glass tube, a much greater tension is to be expected for a given speed. In three experiments, 10-mm diameter glass rods, the ends of which had been polished with rouge and cleaned thoroughly by the procedure previously described, were moved through centrifuged water at a maximum speed of 24 meters/sec.[10] In all three cases, cavitation occurred somewhere in the tube. In one experiment cavitation appeared at the end of the rod when the velocity was 16 meters/sec. In another, an isolated bubble formed below and at a distance from the rod in the first frame, followed by additional bubbles when the velocity was 21 meters/sec., as shown in Fig. 9. The first isolated bubble expanded to a large size and later collapsed so suddenly that the glass tube was broken by the water hammer effect. Cavitation may thus appear at the end of the rod despite the presence of a large bubble cavity in the tube. In the third experiment no cavitation formed on the end of the rod, but after the rod had attained its maximum velocity of 24 meters/sec. and disappeared from the camera field of view, bubbles appeared and expanded into a large cavity, the collapse of which again broke the glass tube.

## DISCUSSION

In the three experiments with a 10-mm diameter glass rod, despite rigorous cleaning, high hydrostatic pressure treatment, and every precaution to remove and protect from hydrophobic dust particles, cavitation occurred at some point in the tube before the maximum velocity of 24 meters per second was reached. The suction of the 10-mm rod in the 16-mm glass tube must have been considerable, as evidenced by the large size of cavity, but no figures for the tension can be given until the problem of calculation is solved. It is possible that in the short time intervals involved, negative pressures of the order of 100 atmospheres are developed. The experiments are described in detail, because of the interesting phenomena involved, which could only be detected by the high speed photographic technique.

In most experiments by other workers on cavitation at the surface of moving bodies, the water has not been freed of gas nuclei and cavitation occurs at a relatively low velocity. It is therefore of considerable interest to find that in the case of experiments with the 5-mm diameter rod moving in a 16-mm tube, the velocity reached

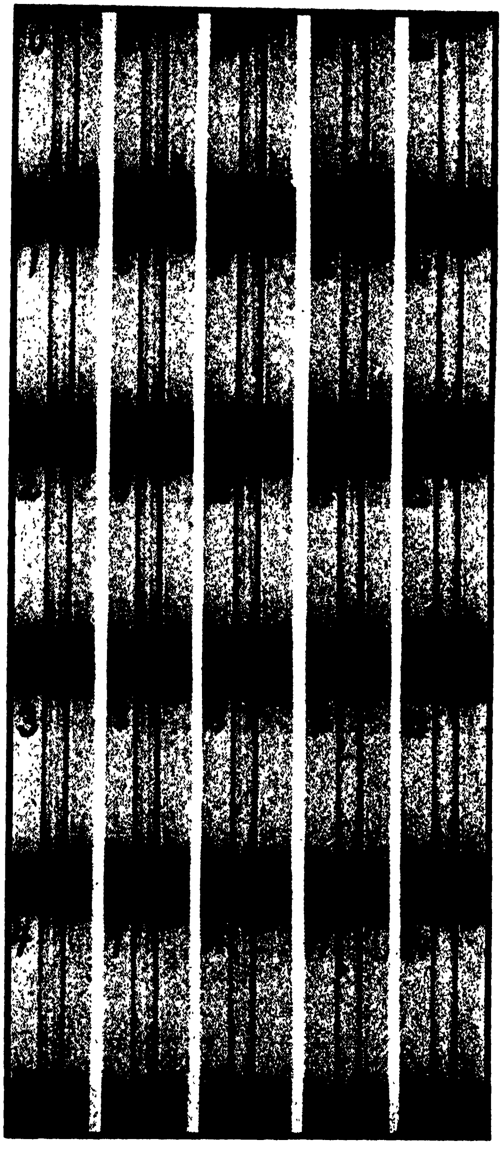


FIG. 7. Moving picture (2160 frames/sec.) of a 5-mm diameter glass rod moving in water at a maximum speed of 16 meters a second. No cavitation appears on the end of the rod until frame 18 but in earlier frames (1, 8, 10) bubbles form away from the rod and grow in size. They probably come from hydrophobic dust particles, as the Cellophane protective covering of the tube was not used in this experiment. Scale in centimeters. Exp. 9-2-44.

---

[10] This rod, which weighed 53.3 grams, when moved with the steel bow in air, attained a maximum speed of 26 meters/sec.

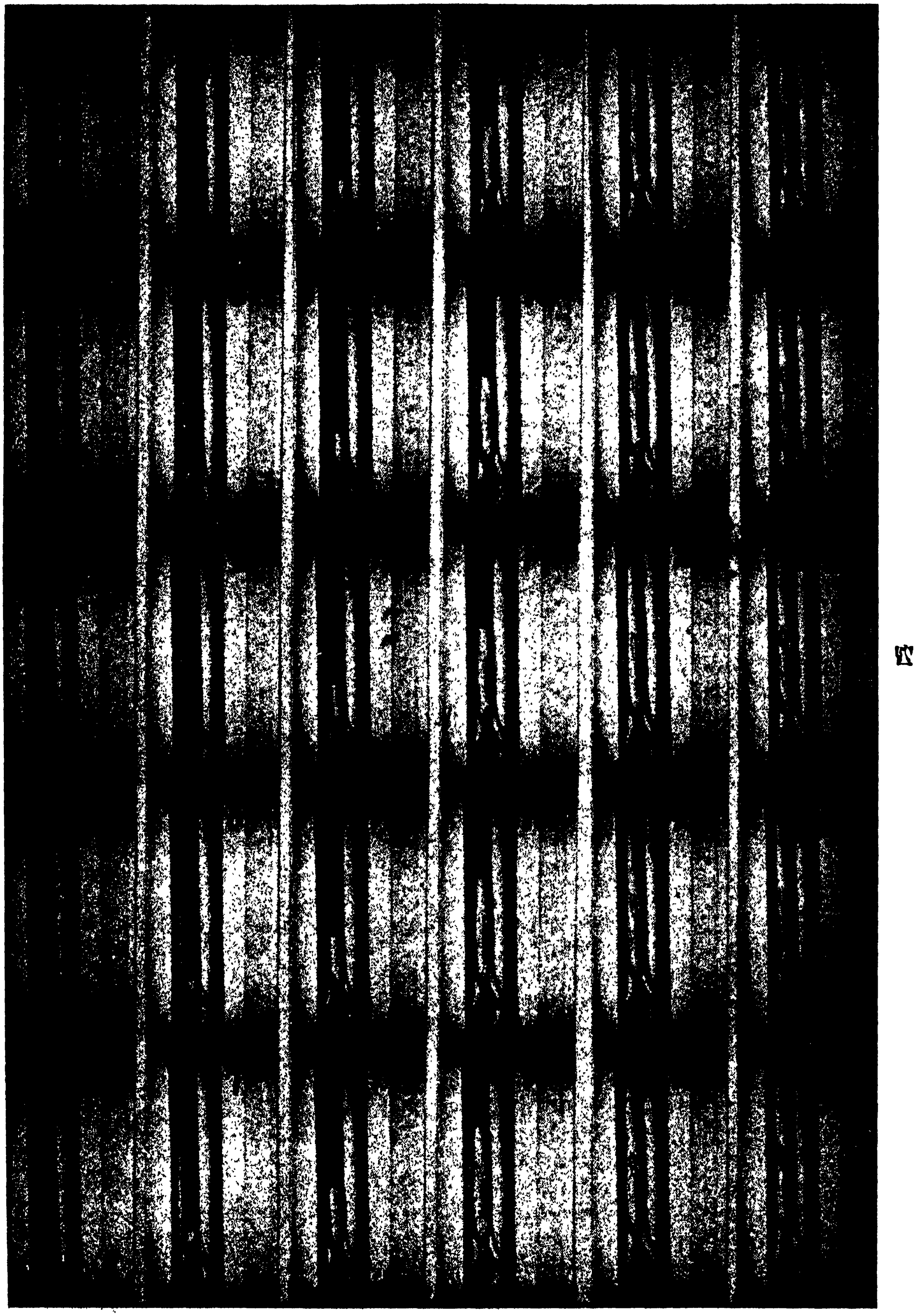
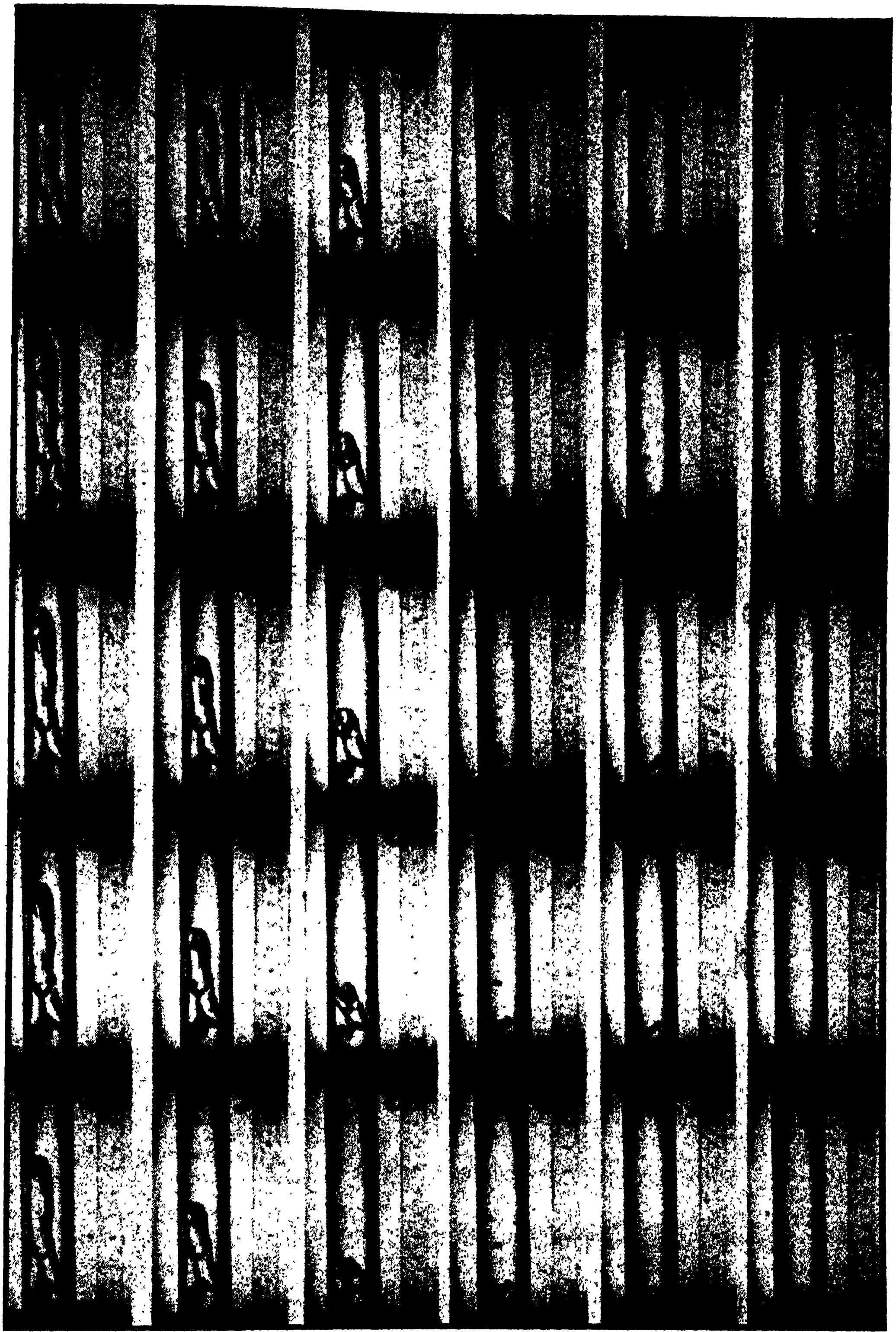


FIG. 8. A and B. Moving pictures of a 5-mm diameter aluminum rod moving in gas nucleus-free corn syrup (viscosity, 20.1 poises; density, 1.383; temperature 23°C) at a maximum speed of 12.2 meters per second. Since the high viscosity of the syrup prevents its filling the space behind the rod, a large



cavity forms from a number of initial points which remain as fingers. In B the collapse of the cavity is shown, beginning with frame 50. Note that collapse is from above downward and that the remnants of the cavity disappear in frame 74 and then reappear again. They oscillate several times before subsiding. Scale in centimeters. Exp. 8-23-44.

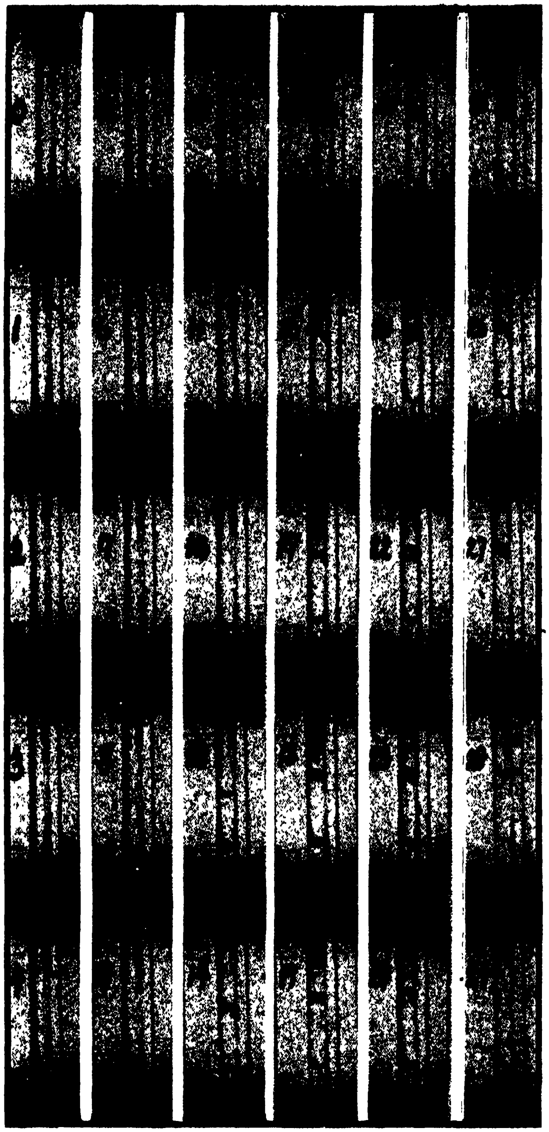


FIG. 9. Moving picture of a 10-mm diameter glass rod, moving in a 16-mm internal diameter tube containing gas nucleus-free water with a maximum speed of 23.2 meters/sec. Note that no cavitation forms on the end of the rod until frame 15 (speed 23.2 meters/sec.) but that a bubble appears below the rod in frame 1 (speed less than 3 meters/sec.) and grows into a large cavity. Bubbles also appear below the rod in frame 13 (speed 21 meters/sec.) and then grow into a mist of cavitation. Collapse of the large cavities broke the glass tube. The two horizontal lines are 10 cm apart. Exp. 12-21-44.

37 meters per second without cavitation appearing anywhere, even at the rear of the rod. This velocity can be better compared with other speeds in water if converted into the English system when it becomes about 83 miles an hour, or 1.38 miles a minute, truly a remarkable figure for cavitationless motion in a liquid medium.

## SUMMARY

Phenomena involving the tensile strength of water have been studied by a kinetic method—high speed motion photography of the rapid movement of a blunt glass rod in a narrow (16-mm inside diameter) glass tube of water.

Special precautions have been taken to remove all hydrophobic patches and small gas masses (gas nuclei) but to retain the dissolved gas (air at one atmosphere) in the water.

If the rod (5-mm diameter) surface contained gas nuclei, or was hydrophobic and free of gas nuclei, cavitation occurred at the rear end when the velocity was less than 3 meters/sec., but if completely hydrophilic and free of gas nuclei, the velocity could be 37 meters/sec. or 83 miles/hour without cavitation.

When improperly cleaned rods (5-mm diameter) were used, isolated cavities formed at intermediate velocities. These cavities appeared and disappeared as the rod movement proceeded, in a manner similar to the periodic cavitation in a Reynolds tube.

Addition of a detergent (diactyl sodium succinate) to the water did not prevent cavitation at a low velocity with the hydrophobic rod (5-mm diameter) free of gas nuclei.

Movement of a rod (5-mm diameter) in pure corn syrup (viscosity 20.1 poises), free of gas nuclei, left a large cylindrical cavity that collapsed in a matter of hundredths of a second. In 75 percent corn syrup (viscosity 0.56 poise) no cavitation occurred.

Large glass rods (10-mm diameter) could not be moved more than 24 meters/sec. within a glass tube (16-mm internal diameter) containing clean gas-nucleus free water without the appearance of cavitation at some place in the liquid, though not necessarily at the end of the rod.

It is not possible to calculate the tensions developed in these "leaky piston" experiments but it may be pointed out that the velocities attained without cavitation are far higher than previously observed for movement of bodies in an aqueous medium, a result believed to be owing to the absence of all gas phases and hydrophobic surfaces.

# A New Magnetic Material of High Permeability

O. L. Boothby and R. M. Bozorth
*Bell Telephone Laboratories, Murray Hill, New Jersey*

(Received November 18, 1946)

This paper describes the preparation, heat treatment, and properties of *supermalloy*, a magnetic alloy of iron, nickel, and molybdenum. In the form of 0.014 in. sheet it has an initial permeability of 50,000 to 150,000, a maximum permeability of 600,000 to 1,200,000, coercive force of 0.002 to 0.005 oersted, and a hysteresis loss of less than 5 ergs/cm$^3$/cycle at $B=5000$. Transformer cores made of insulated 0.001 in. tape, spirally wound, have about the same initial permeability and a maximum permeability of 200,000 to 400,000. The alloy has a Curie point of 400°C and appears to have an order-disorder transformation temperature somewhat above 500°C.

DURING the last few years a magnetic alloy has been made in these laboratories that has an initial permeability several times as high as that of molybdenum permalloy and mumetal, alloys now in general use. The new material has been named "supermalloy" (su-perm′-al-loy). Its preparation, heat treatment, and properties are described briefly in this article. It has already been made in considerable quantities for use in apparatus supplied to the U. S. Navy. When in the form of 0.001 in. insulated tape, used in transformer cores, it has an initial permeability of 50,000 to 120,000; this is to be compared to molybdenum permalloy of the same thickness, having an initial permeability in the range 10,000 to 20,000. The use of supermalloy in communication transformers permits a threefold increase in the range of frequencies transmitted, and a pulse duration three times that heretofore obtained.

Supermalloy constitutes the latest step in the development of the high nickel alloys for use at low inductions. The important previous steps are shown graphically in Fig. 1, where the initial permeability is plotted on a logarithmic scale against the year of discovery.

It is interesting to mention briefly the novelty of each of these steps, regarding them in the light of present day knowledge and theory. The addition of silicon to iron, by Hadfield, neutralized partially the bad effect of the carbon always present as an impurity. The high permeability of the iron-nickel alloys, discovered by Elmen,[1] is attributed primarily to their low magnetostriction, as pointed out by McKeehan.[2] The next improvement, due to rapid cooling or "permalloy treatment," is now believed to be associated with the atomic ordering of iron and nickel atoms in the alloy. The addition[3] of one or more elements such as chromium, molybdenum, and copper, lowered the optimum cooling rate so that it could easily be attained in practice. In Neumann's 1040 alloy[4] the beneficial effect of purification of the material with respect to non-

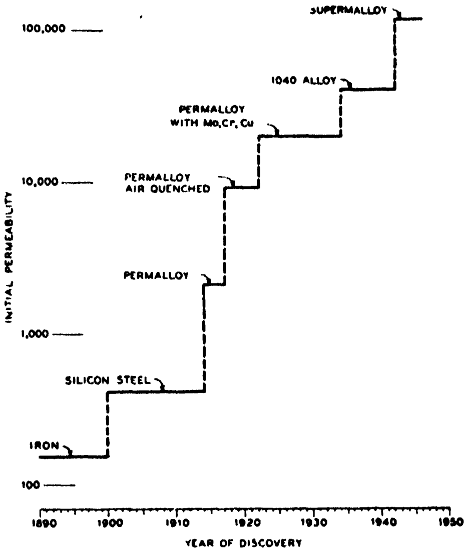


FIG. 1. Initial permeabilities of some useful materials plotted against the year of discovery. Note logarithmic scale.

---

[1] H. D. Arnold and G. W. Elmen, J. Frank. Inst. 195, 621 (1923).

[2] L. W. McKeehan, Phys. Rev. 26, 948 (1925).

[3] G. W. Elmen, Elec. Eng. 54, 1292 (1935); W. F. Randall, J. Inst. Elec. Eng. 80, 647 (1937).

[4] H. Neumann, Arch. Tech. Messen 4, Z913 (1934).

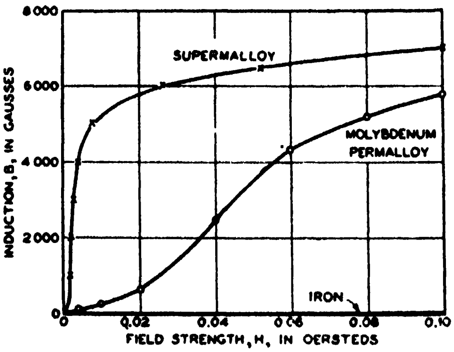


FIG. 2. Magnetization curves for supermalloy, molybdenum permalloy, and iron, in low fields.

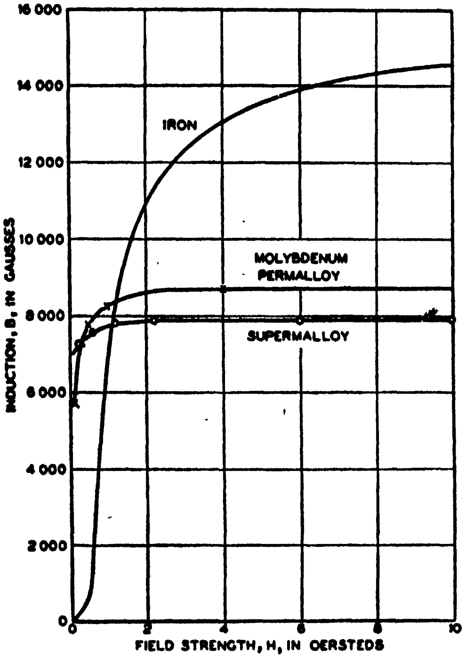


FIG. 3. Magnetization curves in higher fields.

metallic substances such as carbon and oxygen, as developed by Yensen,[5] Cioffi,[6] and others, was also used to advantage.

Although the physical reasons for the success of the supermalloy process are still not known in detail, the work so far accomplished indicates that there are two main considerations: (1) the presence of certain impurities or combinations of impurities, usually found in commercial alloys, will prevent the attainment of high permeability following the procedure now used; and (2) a definite cooling rate must be used below the temperature at which atomic ordering begins, or the specimen must be held for a definite time at a temperature between about 400 and 500°C. The hypothesis is made that, when a critical amount of ordering is present, the magnetostriction and the magnetic crystal anisotropy both tend to disappear at the same time in the alloy of proper composition and that high permeability then occurs in the polycrystalline material. These properties are now under investigation.

## PREPARATION AND HEAT TREATMENT

The composition of supermalloy is about 79 percent nickel, 5 percent molybdenum, 15 percent iron, and 0.5 percent manganese. Impurities such as silicon, carbon, sulfur, etc., are lower than in most commercial alloys. Materials are melted in vacuum in an induction furnace of about 30 lbs. capacity, and poured in helium or nitrogen at atmospheric pressure.

Ingots are hot and cold rolled by commercial methods to any thickness down to 0.00025 in. The tape is wound spirally to form toroidal specimens. When insulation is desired a thin film of magnesia is applied in carbon tetrachloride suspension so that a film about 0.00005 in. in thickness is left on each side of the tape. Trans-

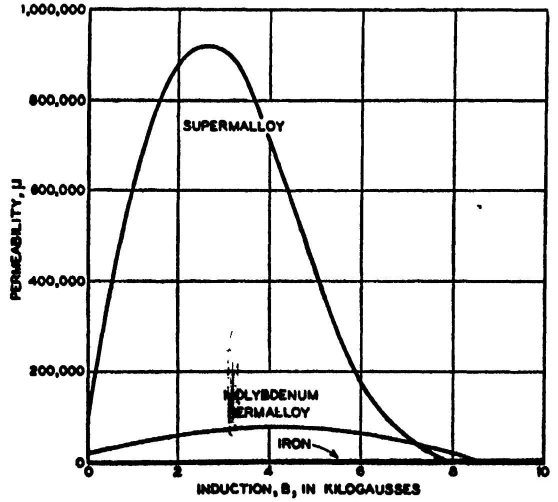


FIG. 4. Permeability *vs.* induction curves of supermalloy in the form of 0.014 in. sheet. Curves for other materials are for comparison.

---

[5] T. D. Yensen, Trans. A.I.E.E. 33, 451 (1914).

[6] P. P. Cioffi, U.S.P. 2, 110, 569 (1938).

former cores are made in this manner. Heat treatment consists of maintaining at 1300°C in pure dry hydrogen, and cooling through the temperature range 600° to 300°C at a critical rate appropriate to the composition.

## PROPERTIES

Magnetic properties are given for (a) 0.014 in. uninsulated and (b) 0.001 in. insulated material. Magnetization curves of the former are compared, in Figs. 2 and 3, with curves of molybdenum permalloy and iron. The ratio of permeabilities of supermalloy and molybdenum permalloy varies with induction from about 5 at $B=20$ to about 10 at $B=3000$ (maximum permeability). At inductions of 7000 to 8000 the permeability of supermalloy has decreased markedly and is less than that of molybdenum permalloy or iron; saturation is at $B_s=7900$. Figure 4 shows typical permeability *vs.* induction curves for supermalloy, molybdenum permalloy, and iron.

The permeability in very low fields is shown in Fig. 5 for a representative specimen of supermalloy, and for materials previously known. Finally, hysteresis loops for supermalloy and molybdenum permalloy are shown in Fig. 6.

When thin tape is *insulated* before heat treatment, the permeability in low fields is affected but slightly. However, the curves of Fig. 7 indicate that the maximum permeability of 0.001 in. insulated tape is considerably lower than that of thicker tape. Insulated 0.004 in. tape has a maximum permeability about twice as large as that of insulated 0.001 in. tape; its permeability is therefore intermediate between the 0.001 in.

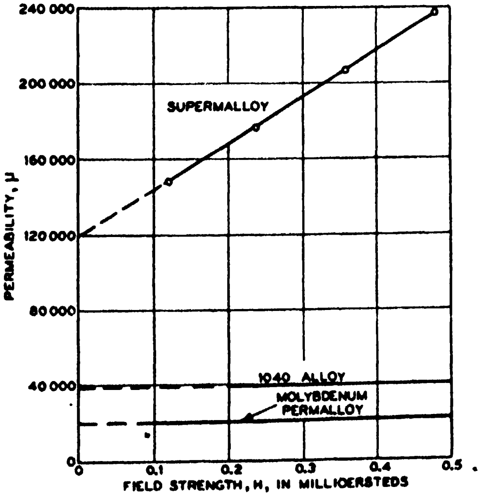


FIG. 5. Permeability *vs.* field strength curves for low fields, for representative specimens of 0.014 in. sheet of three materials of high permeability.

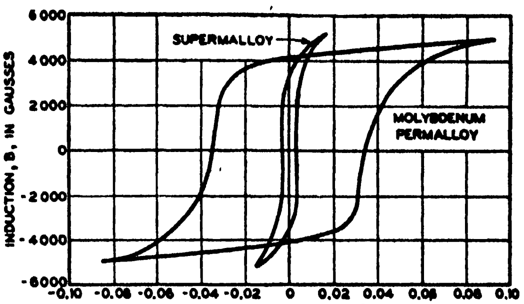


FIG. 6. Hysteresis loops of supermalloy (5 ergs/cm$^3$/cycle) and molybdenum permalloy (50 ergs/cm$^3$/cycle) for $B_m=5000$, in 0.014 in. sheet.

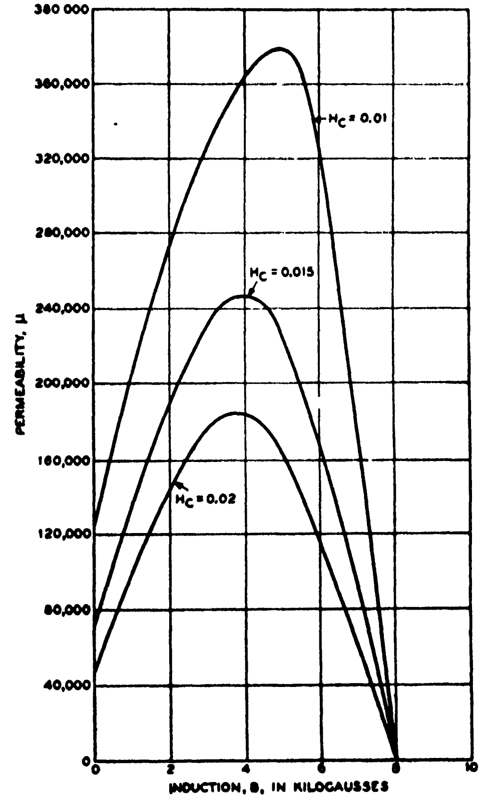


FIG. 7. Representative permeability curves for good, average, and marginal specimens of supermalloy made from insulated 0.001 in. tape.

insulated and the 0.014 in. uninsulated tape. The middle curve of Fig. 7 is near to the average of a lot of several thousand cores of 0.001 in. tape, made in the early stages of development of the material.

The resistivity of supermalloy is about 65 microhm-cm.

The high permeability of supermalloy opens the way toward further improvements in transformers operating at frequencies common in communication engineering, in much the same way that the various permalloys led to the improvements described a decade ago.[7]

We are indebted for cooperation and assistance to many members of the Laboratories and the Western Electric Company. Messrs. D. H. W. Wenny and H. C. Theuerer should be specially mentioned for helping to solve many of the metallurgical problems connected with the development of the material.

---

[7] A. G. Ganz and A. G. Laird, Elec. Eng. **54**, 1367 (1935).

# Apparatus and Techniques for Practical Chemical Identification by X-Ray Diffraction

CHARLES S. SMITH, *Department of Physics, Case School of Applied Science*

AND

RICHARD L. BARRETT,* *Department of Mineralogy and Geology, Case School of Applied Science, Cleveland, Ohio*

(Received August 30, 1946)

The method of chemical identification by means of x-ray diffraction is reviewed from the point of view of apparatus and techniques recently made commonly available, with particular emphasis on the use of long wave-length x-radiation. Experience in using the A.S.T.M. card index file of x-ray diffraction data is reported. The discussion is designed especially for the person who wishes to make use of this important analysis tool but who is not an expert in x-ray diffraction.

## I. INTRODUCTION

WHEN Hull in 1919 described the powder method of x-ray diffraction, he observed that the method should be a potent tool for the identification of crystalline substances. Hull pointed out, "That every crystalline substance gives a pattern; that the same substance always gives the same pattern; and that in a mixture of substances, each produces its pattern independently of the other, so that the photograph obtained with a mixture is the superimposed sum of photographs that would be obtained by exposing each of the components separately for the same length of time."[1] The new method immediately began to be used by mineralogists and chemists but at first its usefulness was limited by the fact that standard patterns of known substances were needed for comparison. In 1938 Hanawalt[2] and some of his co-workers at the Dow Chemical Company published tables giving the diffraction patterns of 1000 substances and showed that the data could be indexed in such a way that the identity of any listed substance could usually be found by reference to the three strongest lines in its diffraction pattern. The availability of this table of standard patterns, including as it did most of the more important inorganic compounds, made x-ray diffraction a practical means of quick identification of an unknown substance. Following Hanawalt's publication the American Society for Testing Materials made available a card index comprising the substances he had listed plus additional ones obtained from other sources and arranged essentially according to the indexing system suggested in his paper. More recently the A.S.T.M. has issued a supplementary set of cards greatly amplifying the scope of the index so that it now includes some 3000 substances. This card index, which makes available in useful form the patterns of the commoner inorganic salts, most of the important minerals, and even many organic substances, enables the x-ray diffraction method to take its place along with the spectroscope and the polarizing microscope among the potent physical tools for chemical research.

In his 1938 paper Hanawalt described in detail the technique used in the Dow Chemical Laboratories and his co-workers and others have amplified the description in subsequent papers.[3-5] The Dow technique employs molybdenum radiation, large radius quadrant cameras, and a multiple diffraction unit which permits the exposure of 24 samples simultaneously. It is perhaps particularly adapted to large industrial laboratories where considerable numbers of samples are being run on a day to day basis. However exposures of some six hours are required and the particular apparatus used is no longer being manufactured.

In the last several years the writers have had occasion to examine a considerable number of

---

* Now at the New Mexico College of Agriculture and Mechanic Arts, State College, New Mexico.

[1] A. W. Hull, J. Am. Chem. Soc. 41, 1168 (1919).

[2] J. D. Hanawalt, H. W. Rinn, and L. K. Frevel, Ind. Eng. Chem. Anal. Ed. 10, 457 (1938).

[3] L. K. Frevel, Ind. Eng. Chem. Anal. Ed. 14, 687 (1942).

[4] L. K. Frevel, Ind. Eng. Chem. Anal. Ed. 16, 209 (1944).

[5] W. P. Davey, J. App. Phys. 10, 820 (1939).

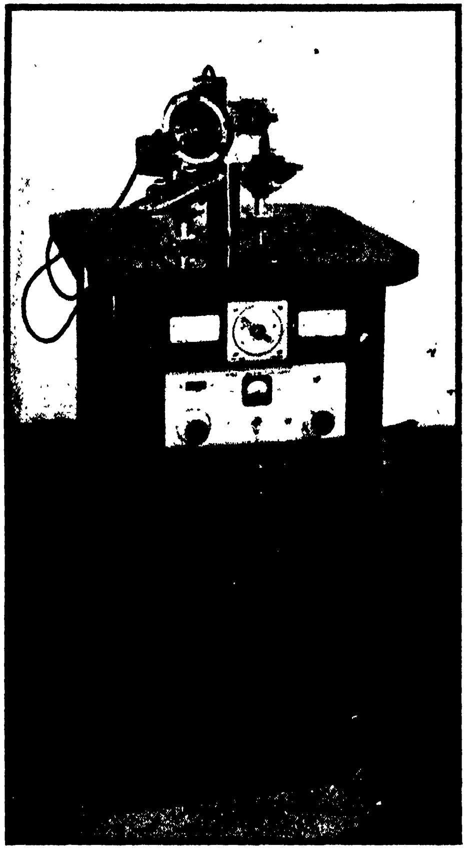


FIG. 1. An example of a modern self-contained x-ray diffraction outfit (Courtesy of the Picker X-Ray Corporation).

industrial samples and, in common with many other workers in x-ray diffraction, have come to prefer a somewhat different technique employing a smaller radius camera of the full circle type, and copper, iron, cobalt, or chromium radiation which makes possible satisfactory resolution of lines with the small camera. Although numerous papers have been published describing details of camera design, special methods of sample preparation, improvements in x-ray tubes, etc., the chemist or mineralogist who is not himself an x-ray specialist has remained dependent for a general treatment of the subject upon the papers of Hanawalt and his co-workers. It is the purpose of the present writers to describe for the nonspecialist the practical considerations involved in modern x-ray methods, employing the more modern apparatus now available commercially. While in a general way the technique to be described will be familiar to x-ray crystallographers some details based upon the writers' experience may be new and it is hoped that the present paper may be helpful to chemists and technologists who have not previously employed x-ray diffraction in taking advantage of this extremely useful tool.

In what follows it will be assumed that the reader has a general knowledge of the physics of x-rays and of x-ray diffraction such as may be obtained in a standard text.[6] It is also assumed that the reader is familiar with the general method of x-ray diffraction chemical identification as presented in the 1938 paper of Hanawalt.

## II. APPARATUS

In the past ten years a number of manufacturers have introduced lines of x-ray diffraction equipment built around the recent American development of efficient diffraction tubes, particularly those with long wave-length targets. In general this equipment is excellent and our intention here is to point out only the practical details in choosing equipment for chemical identification work. The writers' own experience has been in a laboratory primarily devoted to instruction and research and they accordingly have had the benefit of a variety of versatile equipment. They have found it best however to do most chemical identification work with the standard forms of commercial apparatus, employing the available research apparatus only on the rare problems of a difficult or special nature. It is certain that the laboratory doing routine problems will do well to standardize on convenient commercial apparatus chosen for the particular field of identification involved.

### Power Supplies

The available commercial power supplies are designed for use with a particular manufacturer's tube, Fig. 1, and are all uniformly good, incorporating the necessary electrical controls, high voltage, and x-ray protection for the operator and water and electrical failure protection for the apparatus. For identification work alone x-ray tubes of all types can be run at 40 kv and hence the high potential controls are strictly unneces-

---

[6] A new book which presents an excellent treatment is C. S. Barrett, *Structure of Metals* (McGraw-Hill Book Company, Inc., New York, 1943).

sary. Full wave rectification is desirable, and half-wave rectification somewhat less so but neither is necessary. The writers have used for some time a homemade power supply which merits description not because of its originality but because it emphasizes how simple and convenient such equipment can be.

The apparatus employs a ray proofed diffraction tube mounted horizontally on a wide, high wood table. Screens are provided for high voltage protection, but these would be unnecessary were the tube also shockproof and equipped with a shockproof cable. The electrical apparatus consists simply of a high voltage transformer and a filament transformer (both the usual secondhand radiographic equipment), a Variac, millimeter, and voltmeter. The high voltage transformer has a 35-kv secondary and is powered directly from the a.c. line through an ordinary toggle switch which is the only high voltage control. The transformer is located under the table and is protected, the high potential being led by a shielded cable to the tube through the millimeter. The tube filament is powered and controlled by the sequence, a.c. line, Variac, filament transformer. The latter is mounted on the table at the screened cathode end of the tube and supports the filament voltmeter and the milliameter. The cooling water for the tube is carried by Saran tubing, and is run off through a small standpipe in which is immersed a one-ampere fuse which carries the primary current for the high voltage transformer. This water fuse has been effective protection against water or memory failure. The whole outfit is obviously inexpensive, it is simple in its controls, two switches and a Variac; and the horizontal mounting of the tube and the large table area has permitted the easy use of a variety of x-ray cameras.

## Tubes

It has already been mentioned that diffraction apparatus is built around the tube that is used. The diffraction tube is therefore the item on which the choice of apparatus turns and while most features of diffraction tubes, such as a line focus, two or four radiation ports, rated voltage and current, etc., are now standard, there are a few features which vary among tubes and which should be considered in choosing apparatus for chemical identification. Chief among these are window material, effective tube radius, and target material. The first two items determine the x-ray intensity available at the camera, and the choice of the third is influenced by the field of the identification work.

At the present time Lindemann glass and beryllium are used for x-ray tube windows. For Mo target tubes ($\lambda = 0.71$ kX) the difference between these two window materials is not important because the transmission of either is near to one. These window materials were developed for the softer radiation from targets of Cu ($\lambda = 1.54$ kX), Co ($\lambda = 1.79$ kX), Fe ($\lambda = 1.93$ kX) and Cr ($\lambda = 2.29$ kX) where x-ray transmission is inherently low especially with Cr, and here the difference between the two materials is significant, the advantage being with the beryllium.[7]

The effective radius of an x-ray diffraction tube should be as small as possible because one deals always with an inverse square law of radiation intensity. The effective radius is determined by how close one can place the diffraction camera to the center of the tube focal spot. Beryllium window tubes enjoy two advantages here: the camera can be placed in contact with the window if desired, and the window can be placed close to the target, beryllium being a good thermal and electrical conductor.

It has been indicated that the writers have found it advantageous to use characteristic radiations other than the Mo$K\alpha$ recommended by the A.S.T.M. practice.[8] Radiations such as Cr, Fe, Co, and particularly Cu have long been used extensively by crystallographers and metallurgists and now that high intensity tubes with these targets are available in this country, their use in identification problems has become more common. These radiations are all much longer wave-length than Mo, and the choice of the particular one to use will depend strongly on the fluorescence considerations discussed below. Because Cu targets can be safely loaded to higher values than Cr, Fe, and Co targets, Cu-radiation is by far the most commonly used of the group

---

[7] R. R. Machlett, J. App. Phys. 13, 398–401 (1942).

[8] "Tentative Recommended Practice for Identification of Crystalline Materials by Hanawalt X-ray Diffraction Method," A.S.T.M., Philadelphia, Pennsylvania, 1942.

in crystallography and in our own practice. The advantages of long wave-length radiation will therefore be discussed in terms of Cu$K\alpha$; the same remarks will apply to the other radiations with minor changes.

An important advantage of Cu-radiation is the short exposure time required. There are several obvious factors such as necessary camera diameter and film sensitivity which greatly favor Cu-radiation in this respect, but there are also a number of less obvious and equally important factors some of which favor short wave-length radiation. We shall not try to assign a reason therefore but merely state the well known empirical result that for equal dispersion and resolution, a 15-minute Cu exposure is equivalent to one of several hours with Mo-radiation. Such short exposures have made it possible for the writers to supply useful answers to identification problems when the longer exposure would have been prohibitive, and it would seem that workers closer to industrial problems would very frequently find short exposure an advantage.

Since the wave-length of Cu$K\alpha$ is approximately twice that of Mo$K\alpha$ much greater angular dispersion results, the diffraction pattern of most materials being spread through the full 90° range of Bragg angle in contrast to the range of $\theta$ of from 0 to 30° found with Mo$K\alpha$. This fact may be used to help shorten exposures through decreased camera diameter as noted above, or the large dispersion of long wave-length radiation may be used for better observation of those $d$ values lying in the range from 3 to 8 kX. Many materials, especially mineralogical and organic materials, have strong lines, and what is important, discriminating lines, in this region. The writers have found that this $d$ region is indeed a critical one and merits close attention.

In chemical identification work it is seldom necessary to resort to precision determination of $d$ values or lattice parameters. Some specialized problems however are profitably studied by such determinations which are not particularly difficult to make by means of well-known back reflection methods. Back reflection methods necessarily employ long wave-length radiation.

These factors influence us strongly to the use of long wave-length radiation and it is apparent from the availability of the equipment that other workers are so influenced also. There are disadvantages in the use of long wave-length radiation, notably the fluorescence difficulty and the fact that much existing data lists intensities in terms of Mo-radiation. The effect of these difficulties will be discussed, but we are left with a strong preference for long wave-length, and particularly for Cu-radiation, the most generally useful of all.

The fluorescence difficulty may be stated as follows: the characteristic radiation from a given target in the long wave-length group (wave-lengths listed above) is capable of exciting $K$ fluorescence in specimen atoms which have an atomic number less than that of the target by two or more. The resulting $K$ fluorescence radiation ($L$ fluorescence is seldom troublesome) produces a general background on the diffraction pattern which is exceedingly strong if the specimen contains a substantial portion of material located 2, 3, or 4 places below the target material in atomic number. Fluorescence may thus involve an important group of elements, e.g., Co, Fe, and Mn in the case of a Cu target tube, and Mo-radiation possesses the advantage that few important elements are located immediately below Mo in the atomic number.

It is evident then that the choice of the particular long wave-length radiation must be made with the field of the identification work in mind. Thus metallurgists dealing exclusively with ordinary steels use Co- or Fe-radiations; those dealing with high alloy steels would be well advised to use Cr-radiation; chemists concerned only with iron corrosion products should use Co- or Fe-radiation. The person confronted with materials which range the periodic table will find, however, that Cu-radiation is the most generally useful, the number of problems involving Co, Fe, and Mn in such a case being statistically not too numerous, and these problems being susceptible to techniques pointed out below.

The ideal solution in the general case for the fluorescence difficulty is to have available a choice of long wave-length tubes. The writers have fallen back on this expedient upon occasion, but most frequently they actually have found it more convenient to use Cu-radiation and to accept the fluorescence background. The reasons for this procedure may be discussed in terms of Cu-radiation and a specimen known to contain Fe. Four cases arise depending on whether the material is single phase or polyphase, and on whether the phases are simple or complex crystallographically. A single simple phase such as Fe or FeO being high in Fe will produce strong fluorescence when exposed to Cu-radiation, but being simple, it will in general produce a strong diffraction pattern, and we have found such patterns easily identifiable in spite of the background. The same remark applies to a lesser degree to a polyphase mixture of simple materials. Single phase specimens involving crystallographically complex materials will in general give a weak diffraction pattern, but also in general the proportion of Fe and hence the fluorescence will be low. Polyphase and crystallographically complex materials will usually be difficult, and so, of course, would be the detection of any minor phase in the presence of a high Fe major phase. In the Cu-Fe situation and in similar situations the fluorescence background can be reduced by the use of filters discussed below, but this technique is not nearly so effective as when the fluorescence is caused by elements considerably below the target element in atomic number.

Elements five places or more below the target element also produce fluorescence, but in this case the excitation is not nearly so strong, and since the fluorescence wave-length is considerably longer than that of the target radiation, it can be effectively filtered out of the radiation striking the film. Filtration can be effected by placing between the sample and the film sheets of aluminum or better still, sheets of the $K\beta$ filter appropriate to the radiation used. In this manner identifiable patterns of $MnO_2$, and excellent patterns of $TiO_2$ can be made with Cu-radiation. Because fluorescence radiation may be excited by radiation from the tube other than the characteristic radiation which creates the diffraction pattern, the use of the $K\beta$ filter in this fashion is generally desirable and should be standard practice. Since the filter used in this way should not be in contact with the film and because filter materials are difficult to handle, some sort of jig or other holding device is desirable. It would be helpful if camera manufacturers built in such devices in the future.

### Filters

The purpose of the $K\beta$ filter is to reduce to a negligible amount the intensity of the characteristic and monochromatic $K\beta$-radiation relative to the intensity of the $K\alpha$-radiation of the target element. In the case of the long wave-length targets this can be done by inserting in the x-ray beam a filter composed predominantly of the element one lower in atomic number than the target element; for Mo-radiation a proper filter is Zr. Such filters may be the proper pure element and in the case of Ni for Cu-radiation this is convenient[9] but in the case of other $K\beta$ filters the active element is more easily used in the form of a compound. Such filters made with the proper concentration of active elements are now available in any desired size.[10] The filter will serve its designed purpose when a small piece is inserted in the primary x-ray beam, but as indicated above, the $K\beta$ filter will in addition act as a fluorescence filter when introduced in the proper sized sheet between the film and the specimen. Since the $K\beta$ filter does not eliminate the $K\beta$-radiation but merely reduces it relative to the $K\alpha$-radiation, the identification worker will do well to assess its effectiveness by making long exposure trial patterns of a simple substance.

### Camera Design

At the Dow Chemical Laboratories quarter-circle cameras of 8-inch radius are used on an apparatus which permits exposing 24 samples at one time. However the exposure time runs about six hours. Where molybdenum radiation is used, the quarter-circle camera gives all of the pattern that is useful for identification purposes. With longer wave-length radiation, it is profitable to employ a full circle camera which permits recording of back reflection lines and gives a symmetrical pattern which doubles the accuracy of measurement of the angle $\theta$.

Cameras of the full circle type are commerci-

---

[9] Nickel foil of the proper thickness (0.0004 in.) may be obtained from the General Electric X-ray Corporation.

[10] Manufactured by the Patterson Screen Company, Towanda, Pennsylvania, and available through representatives of the leading x-ray apparatus manufacturers.

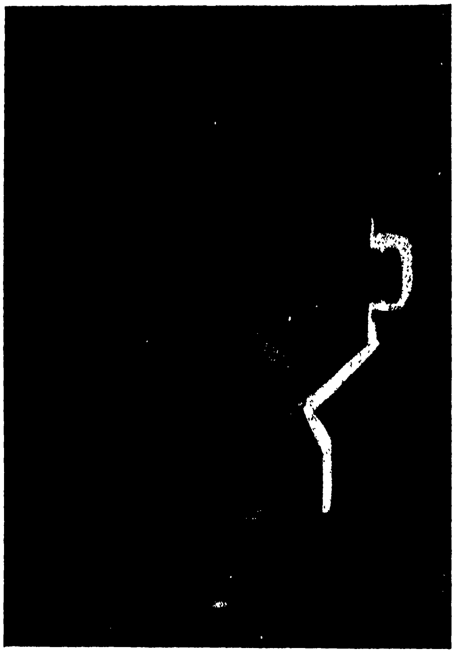


FIG. 2. An example of an x-ray diffraction camera particularly suitable for identification work. The collimating system is entirely within the camera. Two desirable features shown are the specimen centering device and the primary x-ray beam outlet shield (Courtesy of Otto von der Heyde, Newton Highlands, Massachusetts).

ally available from several of the manufacturers of x-ray equipment and also from some of the independent instrument makers. These cameras differ among themselves in a number of details, and the purchaser of new x-ray equipment will do well to investigate a number of makes with a view to getting the most suitable equipment for his purpose. The design of powder cameras for general use has been extensively treated recently in an excellent paper in this journal.[11] We shall mention here points particularly pertinent in chemical identification.

The diameter of the camera should preferably be some integer multiple of 57.26 mm in order that linear measurement on the film will be easily convertible to angular measurement in degrees. This greatly reduces the amount of routine calculation in determining *d* values. The writers have used a camera of 57.26-mm diameter with considerable satisfaction and have found it to give sufficient accuracy for most routine identification work. With copper radiation, exposures run from 10 to 20 minutes. However, we are of the opinion that if only one camera is to be purchased and speed is not too important a consideration, it would be best to select one of 114.52-mm diameter. This camera size used with copper radiation permits somewhat greater accuracy of measurement than the large radius quadrant cameras employed by Hanawalt when used with molybdenum radiation. By using the back reflection lines, lattice parameters for cubic crystals can be calculated with considerable precision. At the same time exposure times are not unduly large. Cameras of still greater size may be useful on occasion, but should not be needed for routine identification work.

Some provision for rotating the sample during exposure is made on practically all modern cameras. The purpose of sample rotation is to increase effectively the randomicity of crystalline orientations in the specimen. In the case of large grained, difficult-to-grind specimens, rotation saves much time and labor. The writers prefer the type of apparatus for rotation with a belt driven pulley attached to the sample holder rather than the built-in motor type. The reason is that the sample can be rotated by hand while the alignment of the x-ray beam is being adjusted, thus checking the centering of the sample. We prefer not to rotate the sample if the sample size is adequate, and the particle size small enough to give a smooth pattern without rotation. It is very important that the sample should be accurately centered in the camera and coaxial with the rotating mechanism. While a number of permanently centered cameras are on the market, the type with centering screws for making this adjustment is to be preferred.

All modern cameras have a fluorescent window in the exit portal for use in aligning the camera in the x-ray beam, and the centering of the sample can be finally checked by watching the shadow of the sample in the window while the sample is rotated by hand. The design of the collimating system is of utmost importance in obtaining clean films with a minimum of background darkening. It is desirable for intensity reasons that the collimating system should be within the camera itself with the final opening as near as possible to the sample. Some cameras have as a highly desirable feature a shield or trap

---

[11] M. J. Buerger, J. App. Phys. 16, 501–510 (1945).

which protects the film from radiation scattered from the collimating system and from the air near the exit portal, Fig. 2.

The camera must be provided with a flexible and convenient mounting arrangement so that it can be easily aligned in the x-ray beam. It should also be possible to remove the camera for loading and unloading films and replace it in accurate alignment without readjustment. A little time saved on these routine operations will mount up if much work is done.

### Film

At the present time x-ray film is available with thick, double emulsions, which film compares favorably in speed with older emulsions used with intensifying screens. Such film sold variously under the names No-Screen, Non-Screen, etc., is best for diffraction work. The writers always use this type of film and in routine work make no attempt to correct for dimensional changes caused by the film processing, finding it readily possible to achieve the accuracy inherent in the A.S.T.M. data without correction. The only precaution taken is to dry the film thoroughly before measurement.

It is proper here to mention the Straumanis technique which has come into common use. The Straumanis technique is essentially a method of eliminating errors owing to camera diameter uncertainty and film dimension changes[12] by an ingenious method of loading the film. Two holes are punched in the film instead of the customary one hole for the exit beam. The two holes are spaced a camera half-circumference apart and the film is loaded so that the beam enters through one hole and leaves through the other. If back reflection lines are present on the pattern, appropriate measurements on the processed film will serve to determine the effective camera diameter. This scheme will fulfill the purpose for which it was designed only when properly used. The technique inherently demands precision positioning of the specimen and, therefore, requires an accurately made camera equipped with specimen centering cross motions. The Straumanis method of loading the film is to be recommended regardless of whether the diameter determination is carried through. No extra effort is involved, and as has been pointed out[11] a complete record of the diffraction pattern results, which can at any latter time be used for diameter correction and/or back reflection measurements.

### Geiger Counter Method

There has recently appeared an excellent Geiger counter spectrometer[13] which has valuable application in the field of x-ray diffraction chemical identification. Because the instrument yields directly and quickly x-ray intensities, it opens wide the field of quantitative diffraction analysis, and it would appear to be especially valuable in routine control operations. While the instrument possesses ample angular precision for identification purposes, the writers are not convinced by the literature statements that the instrument "is capable of angular measurements and resolutions beyond that attainable by means of photographic cameras."[14] The published intensity curves and comparable microphotometer records do not support the part about angular measurements, and it is well known that microphotometer traces are an exceedingly poor way of exhibiting photographic resolution and the detectability of a diffraction line. The instrument will have many applications but we feel that these will be complementary to photographic methods in the general identification laboratory.

## III. TECHNIQUE

### Preparation of the Sample

Probably there are almost as many ways of preparing the diffraction sample as there are x-ray workers. The ideal sample would consist wholly of the substance to be identified without either a container or a binder. The specimen should be small enough for good resolution and to avoid absorption effects and yet large enough for reasonable exposures and for practicable handling. Finally the specimen must contain enough crystals to give a smooth line pattern. These ideals can seldom all be attained simultaneously in practice except in the case of metals which can be drawn into wires.

---

[12] M. J. Buerger, *X-Ray Crystallography* (John Wiley and Sons, Inc., London, 1942), pp. 394–396.

[13] North American Philips Company, Inc., 100 East 42 Street, New York 17, New York.

[14] H. Friedman, Electronics (April 1945).

The size of the specimen to be exposed to the x-ray beam is a matter of compromise primarily between x-ray line resolution and practicable handling. Consensus of opinion among x-ray workers seems to favor a sample diameter of 0.4 mm. For the small 57.3-mm diameter camera, this is certainly an upper limit for good resolution, and the writers have found it necessary to employ smaller sample diameters in many cases.

Specimens which are the proper size but which absorb x-rays highly will produce systematic small errors in $d$ value and large errors in intensity. The cure is reduction in specimen size, error correction, or dilution. Most sample preparation techniques involve a binder which allows qualitative dilution to be made. The writers have never found it practical to follow a schedule of dilution, although this is possible.

Samples should be ground to 200-mesh or finer to help insure that the sample contains enough crystals to produce a smooth line diffraction pattern. Many identification samples have small enough crystal size naturally but must be ground to insure homogeneity and for effective mixing with the binder. Small "sillimanite" mortars are very convenient for grinding and mixing the identification specimen which is often itself quite tiny. Such mortars are so inexpensive that a number can be kept at hand to avoid the necessity of repeated washing, and even to store the specimen for other tests.

Frequently 200-mesh specimens do not produce smooth diffraction lines. In this case the specimen must be rotated, and in some cases should be scanned along its length in order to increase the effective number of crystals. Buerger[11] has discussed the design of rotation devices and of an ingenious scanning device. The latter, unfortunately, is not available commercially to our knowledge. The point to be kept in mind about rotation is that it demands accurate specimen centering; scanning imposes such alignment demands as to be of questionable practicability in routine identification work. The writers prefer not to rotate if it can be avoided by further grinding. An interesting sidelight to rotation is that occasionally in identification work the lines of two phases may be separated by their differing spottiness.

In the Dow Chemical Laboratories, samples are loaded in thin Pyrex capillary tubes having an internal diameter of 0.40 mm. Where molybdenum radiation is employed, glass capillaries are not greatly objectionable, but even very thin glass is quite absorptive of x-rays and most workers who use softer radiations, such as copper or iron, must avoid it. Plastic capillaries have been used by some workers but they are not readily available at the present time. It is quite possible to prepare suitable tubes of Cellophane by rolling it around a mandrel of proper size.[15] The tubes are made as short as the sample holder will permit and the ends are closed with wax after filling. Other workers prefer to mix the sample with collodion diluted with amyl acetate and ether and coat it on a thin thread obtained by untwisting a piece of sewing thread. Another method is to place the sample in a drop of collodion on a glass slide and when the collodion is nearly dry to roll it into the shape of a thin rod incorporating the sample.[16] If only a little sample is available, all of it can be quite conveniently concentrated in one end of the rod. Rubber cement can be used in the same way. Another technique is to mix the sample with a suitable binder such as library paste, dilute collodion, or glyceryl phthalate and then extrude the mixture through a simple die.[17] The die itself with the attached extrusion can be made to fit directly in the sample holder. Probably the beginner in x-ray work will do well to try several of these methods and then adopt the one that pleases him best. It must be remembered that all tubes or binder materials absorb some x-rays and produce more or less background darkening on film, and many of them produce a diffraction pattern of their own. It is desirable to run a blank film on the container or binder without any sample in order to determine whether any diffraction pattern is produced.

### Measurement of Films

After the exposed film has been developed and dried, it may be conveniently studied on a light box with a sloping opal glass front. Some workers use more or less elaborate film measuring devices, but the writers prefer to measure routine films

---

[15] A. A. Burr, Rev. Sci. Inst. 13, 127 (1942).

[16] G. A. Harcourt, Am. Mineralogist 27, 63 (1942).

[17] J. S. Lukesh, Rev. Sci. Inst. 11, 200 (1940).

made in the 57.26-mm diameter camera by setting a pair of sharp pointed dividers on corresponding lines and reading the distance on a good celluloid scale graduated in half-millimeters. Successive readings seldom differ by more than one or two tenths of a millimeter, and errors as large as this are likely to occur in setting any of the measuring devices in use.

When very weak lines must be found and measured, as is sometimes necessary in polyphase samples, the characteristics of the human eye play an important part.[18] The cross line slider with which some measuring devices are equipped has a definite tendency to cause weak lines to vanish, apparently, as the setting is being made. The simple divider technique avoids this difficulty and the detection of weak lines can be made easier by providing the viewing box with a Variac to lower the level of illumination. It is sometimes an advantage to clip or tape the film to a piece of plain glass which can be placed some distance from the opal illuminator and which can be tilted with respect to it. If a device with a slider is used, we prefer a pointed indicator to the cross line type.

If the camera diameter is some simple multiple of 57.26 mm, interlinear measurements are easily converted into angular measurement in degrees by simply dividing by the appropriate factor. $d$ values are calculated from the Bragg formula $d=\lambda/2\sin\theta$. Repeated performance of this routine calculation is time consuming and can be avoided by the use of a very convenient graphical chart prepared at the suggestion of the writers by O. E. Brown.[19] This chart is intended particularly for identification work and covers a range suitable for use with the A.S.T.M. tables. The chart covers copper, cobalt, iron, chromium, and molybdenum radiations which are the ones most valuable for identification work. To use it, the measured $\theta$ value is found under the appropriate horizontal line and the corresponding $d$ value read off above the line.

The intensity of x-ray diffraction lines is most easily estimated by eye and noted on a scale of 1 to 10. Apparently most workers use such a method as do the writers. The method is recommended in the majority of problems for its speed, but some problems may require a more careful method. Hanawalt's method is to use a comparison strip, which is a film containing lines of graded exposure. The comparison method is certainly accurate enough for all identification work, and once the comparison strip is prepared the use of it is nearly as rapid as simple estimation. It is to be hoped that one of the x-ray manufacturers will make comparison strips available commercially.

## IV. INTERPRETATION

The result of the purely experimental side of x-ray chemical identification is a table of $d$ values and the corresponding relative intensities. The interpretation of these data may follow several courses, depending on the complexity of the problem and the comparison data available. Identification work has long been carried out by simple comparison of the unknown film with films of known substances, and much work can be and is done in this way even yet. Such a method fails in the case of polyphase materials or where only a limited library of standards is available. The great contribution of Hanawalt and the A.S.T.M. has been to place in the literature a volume of identified data arranged in a systematic fashion.

### Examples

A great volume of identification problems deal with questions such as "What manganese compound is present in this sample?," "What oxide of lead is this?," or "This substance is predominately one phase; what is the phase?" Where the problem is a simple one of this type, with auxiliary information available, the data for 1000 common materials published by Hanawalt[2] are the rapid and effective means of making the identification. An example is described below in which this interpretation technique was used, and in which also a modified experimental approach was used.

On one occasion the writers were asked to identify a very thin white coating on the inside of a glass tube. The conditions under which the coating had been formed suggested the probability that it was a copper compound. The

---

[18] It may be pointed out here that the eye is a far more sensitive instrument for such purposes than a microphotometer.

[19] Orley E. Brown, J. App. Phys. 18, 191 (1947).

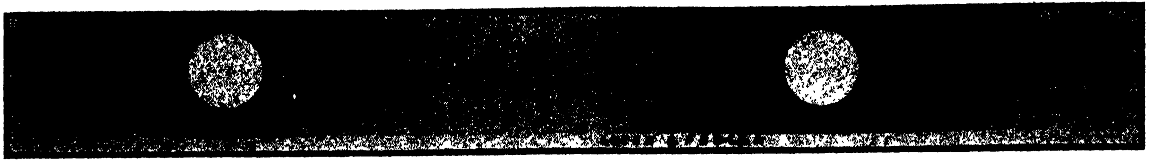


FIG. 3. X-ray diffraction pattern made with Cu$K\alpha$-radiation of an unknown identified as $CaSO_4 \cdot \frac{1}{2}H_2O$ and $CaCO_3$. The line labeled *B* is a superposition of strong lines from each phase. The $K\beta$-filter (a sheet of nickel foil) was placed between the sample and the film. The sharply defined blackening occurs where the filter did not meet and is caused by calcium fluorescence radiation.

thinness of the coating made it infeasible to prepare a sample in the ordinary way. Instead, the tube was broken, and a small piece of the glass was mounted in the center of the camera in such a way that a pencil of x-rays struck the coated surface with a very small glancing angle. A pattern consisting of only four rather weak lines was obtained. The *d* values were computed and the data of Hanawalt[2] were inspected, the search being confined to the simple copper compounds. The data matched perfectly with the principal lines of the pattern for $CuCl_2$. Since the presence of chlorine in the system had not previously been suspected, the information proved highly useful and the determination required less than an hour of elapsed time. Other modifications in technique will be suggested by the ingenuity and experience of the worker. The small 57-mm camera is ideal for problems of this general type, where precision and resolving power are not particularly needed, and where speed may be useful.

The more complex problems involve polyphase samples, or samples consisting of an uncommon single phase. In the former case the interpretation approach described above is often effective with the dominant phase, but the minor phases which may be represented by only a few lines each must be attacked by using a systematic method employing as large a volume of standard data as possible. The A.S.T.M. card index file of x-ray diffraction data provides such a method.

The A.S.T.M. card index file, including the recent supplement, contains data for roughly 3000 substances. A given substance is represented by three cards, one for each of its three strongest lines and their relative intensity, and the card for the strongest line carries, in addition, the data for the entire pattern plus auxiliary information and references. All cards are arranged in order of decreasing *d* value.[20] To use the card file one simply searches the cards in the vicinity of the *d* value of the strongest unknown line, allowing reasonable limits of error. One looks for a card that shows *other* strong lines matching other strong unknown lines, and which corresponds with a substance which is reasonable on the grounds of the other evidence available. When a match of strongest lines is found, the card corresponding with the strongest line of the tentative identification is looked up and all the unknown lines are compared. If all unknown lines match, the identification is complete; if the match is not good, a further search must be made; if the unknown accounts substantially for all the lines of the tentative identification and unknown lines are left over, a second phase must be sought for. The complex situation involving several phases and superposition of lines is best described by example.

Such an example is one which the writers invented for students in an x-ray laboratory class to solve, and the reasoning used by one of the better students is presented below. The sample consisted of equal parts by weight of $CaSO_4 \cdot \frac{1}{2}H_2O$ and precipitated $CaCO_3$, and the students were given no advance information except that of their eyes telling them that the unknown was a white fluffy powder.

The sample was prepared by grinding in a small "sillimanite" mortar, mixing with collodion-ether-amyl acetate and coating on a fine cotton thread. The film was loaded in the Straumanis manner and a filter of nickel foil 0.0004 in. thick was placed in the camera between the sample and the film. In this particular instance a 72-mm camera manufactured by the Picker X-Ray

---

[20] This statement applies to the cards as furnished by the A.S.T.M. Other arrangements within subgroups of cards are possible.

Corporation was used and the exposure was 30 minutes with copper radiation and without rotation of the sample. The resulting pattern, Fig. 3, contained one very strong line, about ten medium lines, and many weak lines. The lines were measured with dividers and scale, the camera diameter was determined from the film, and the $\theta$ values were computed. $d$ values were determined by the use of O. E. Brown's chart.[19] Intensity values were estimated visually. The resulting data are listed in Table I.

Since the line at $d=3.02$ kX is by far the strongest line on the film the student began his attempted identification by searching in the card file from $d=3.00$ to $d=3.04$ kX for cards which also listed a strong line at either $d=5.9\pm0.1$ kX *or* $d=2.80\pm0.02$ kX. When such a card was found, the third listed line was checked against the data for the unknown sample. Those cards on which all three strongest lines did not correspond to lines in the data for the unknown (three cases) were returned to the file. The one case where correspondence occurred for all three lines was the card for $CaSO_4\cdot\frac{1}{2}H_2O$. Accepting this as a tentative identification, all of the lines listed for this substance were compared with the data for the unknown, the cases of correspondence being shown in Table I. It will be seen that the strongest five lines listed for $CaSO_4\cdot\frac{1}{2}H_2O$ as well as some of the weaker lines occur in the data for the unknown. The identification of $CaSO_4\cdot\frac{1}{2}H_2O$ is thus confirmed, but it was clear that an additional phase must be present since so many lines in the unknown were not accounted for. This being the case, the fact that some of the weaker $CaSO_4\cdot\frac{1}{2}H_2O$ lines were missing from the unknown was correctly ascribed to the effects of dilution.

The next step was to identify the as yet unknown phase or phases. At this point the student reasoned that although the line $d=3.02$ is by far the strongest line on the film, it is listed as only the second strongest in the data for $CaSO_4\cdot\frac{1}{2}H_2O$ and consequently there must be a strong line for the unknown phase superimposed upon it. Accordingly he again searched the file in the $d=3.02$ region, this time seeking cards which list a strong line at $d=2.48\pm0.02$ *or* $d=2.28\pm0.02$. This search led to the $CaCO_3$ card which listed as its third strongest line $d=2.28$ and as its second strongest $d=1.92$ which is also found in the data for the unknown. The comparison is completed in Table I which exhibits the fact that all of the experimental $d$ values are accounted for by the identified phases. Conversely the experimental data account for all the strong and medium lines listed in the card file for the two phases except $d=2.13$ for $CaSO_4\cdot\frac{1}{2}H_2O$. Examination of the film revealed that the $d=2.09$ line is in fact broadened toward the high $d$ side suggesting that the $d=2.13$ line was missed because of partial superposition. The intensity agreement is rough but satisfactory. Any doubts on this score are resolved when reference is made to cards which give duplicating data for the two phases from different sources. The experimental data agree with any one card as well as the cards agree with each other.

This example was chosen intentionally to present the student with the difficulty caused by superposition of strong lines from the two phases contained in the sample. Yet the identification was correctly made in a minimum of time. The fact that all the stronger lines for both phases appear on the film would have justified the further inference that both phases are present in major amount. Any further quantitative inferences would not be justified without prepa-

TABLE I. X-ray diffraction data for an unknown specimen compared with the card file data of the identified phases.

| Experimental data | | | Card No. 2438 $CaSO_4\cdot\frac{1}{2}H_2O$ | | Card No. 2438 $CaCO_3$ | |
|---|---|---|---|---|---|---|
| Line | Relative intensity (estimated) | $d$ value in kX | $d$ value | Relative intensity (measured) | $d$ value | Relative intensity (measured) |
| | 4 | 5.96 | 6.0 | 0.40 | | |
| | 3 | 3.87 | | | 3.86 | 0.08 |
| *A* | 3 | 3.46 | 3.48 | 0.30 | | |
| *B* | 10 | 3.02 | 3.00 | 0.60 | 3.04 | 1.00 |
| *C* | 4 | 2.80 | 2.80 | 1.00 | | |
| *D* | 4 | 2.48 | | | 2.49 | 0.20 |
| | | | 2.34 | 0.02 | | |
| *E* | 4 | 2.28 | | | 2.28 | 0.24 |
| | | | 2.13 | 0.18 | | |
| *F* | 4 | 2.09 | | | 2.09 | 0.20 |
| | 3 | 1.92 | | | 1.92 | 0.32 |
| *G* | 3 | 1.86 | | | 1.87 | 0.24 |
| | 3 | 1.84 | 1.85 | 0.60 | | |
| | | | 1.74 | 0.02 | | |
| *H* | 1 | 1.68 | 1.69 | 0.10 | | |
| *I* | 2 | 1.595 | | | 1.60 | 0.16 |
| | | | 1.53 | 0.04 | | |
| *J* | 2 | 1.514 | | | 1.51 | 0.12 |
| *K* | 1 | 1.469 | 1.470 | 0.02 | 1.475 | 0.05 |
| | | | 1.445 | 0.02 | | |
| *L* | 2 | 1.432 | | | 1.439 | 0.08 |

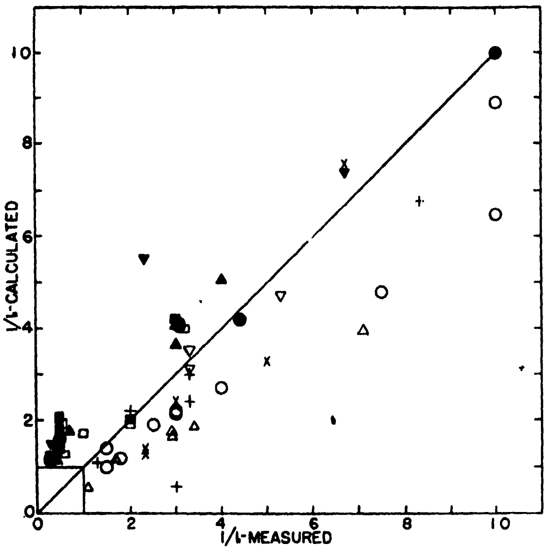


FIG. 4. Calculated theoretical relative intensities are plotted against relative intensities measured by the comparison method by Hanawalt and co-workers. The experimental data all come from the A.S.T.M. card file. The numerous data in the small square are not plotted.

□ Columbium
○ Cadmium oxide
△ Tungsten
▽ Copper
+ Sodium chloride
■ Tantalum
● Cuprous oxide
▲ Aluminum
▼ Lithium fluoride
× Calcium fluoride

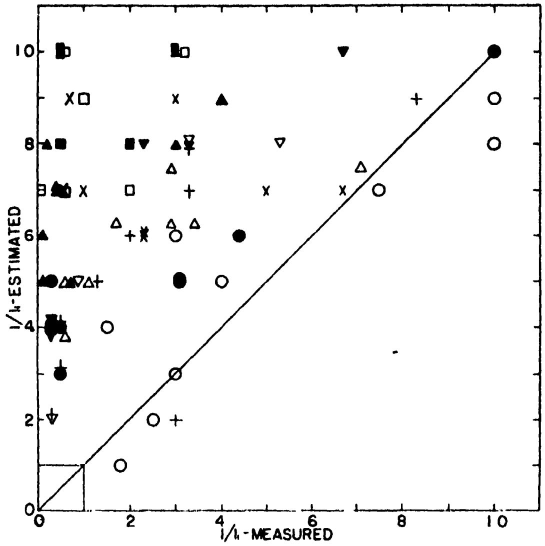


FIG. 5. Estimated relative intensities are plotted against measured intensities for the same substances as Fig. 4. Data are from duplicate entries in the A.S.T.M. card file.

ration of standard films from the two substances for comparison.

### V. AGREEMENT OF DATA

The Hanawalt scheme of chemical identification is based on the empirically demonstrated fact that the *d* values of the three strongest lines of a phase identify that phase nearly uniquely. Completely unique identification implies agreement of all *d* values and in addition substantial agreement of intensities. While *d* value agreement, within allowable limits, must occur intensity agreement often does not occur even in the case of the three strongest lines. Often, too, different sources will disagree on the actual presence of a fairly important line.

Table I lists three sets of data for the same substance, taken from the A.S.T.M. card file, and illustrating these three effects. These data are the work of skilled workers, and there is little likelihood of chemical impurity of the sample examined. It is highly probable that the reader will find as many deviations in his own data as among the ones quoted. The agreement among *d* values is about as good as could be expected considering technical limitations. *d* value results similar to the third set of data quoted *may* occur because of excessive sample size and absorption, poor sample position or possibly film shrinkage, all effects which should be eliminated in standardized technique. The situation with regard to "missing" lines is an interesting one of not infrequent occurrence, the reasons for which are too varied for a general explanation to be attempted here. The very existence of such discrepancies in the card file means that one need not be too disconcerted when it occurs in practice. Of course, a reason and sometimes a significant reason for such effects exists, but if the work has been carefully done the "difficulty" may just as well be in the literature data or literature material.

The discrepancies in intensities of Table II are disturbing at first sight and the writers have made an inspection of published data in an attempt to discover the source of the difficulty. The details of the study need not be given here but the results of the comparison will be of interest to the user of the card file data and are summarized in Figs. 4 and 5.

Duplicate sets of data for 10 simple substances are compared in the figures. Since the chosen substances are simple, reliable calculations of theoretical intensity are readily made by the accepted methods. Figure 4 is a composite plot comparing the relative intensities as measured by Hanawalt by his simple method of visual comparison with prepared standards, against the calculated theoretical intensities. The calculated and measured values agree as well as is usually the case with x-ray diffraction intensities. The comparison is presented here to give added reliance on the measured values and to indicate the degree of intensity agreement to be expected generally.

Figure 5 gives a comparison of intensities as measured by Hanawalt with intensities obtained by the commonly employed method of visual estimation on the basis of an arbitrary scale of one to ten. The data are all from the card file and the number of observers is large enough to eliminate the personal factor. It is clear that the estimated intensities differ rather radically from the measured ones and that the method of estimation tends to exaggerate the relative intensity of the weaker lines. The disagreement is large enough that an identification cannot be expected to show intensity agreement except in a highly qualitative fashion when intensities are estimated. Actually in straight identification work, the disagreements are not as important as they appear. The intensity values themselves are less important than the *order* of intensities and this is quite reliably determined by visual comparison.

Since one of the purposes of the present paper is to recommend the use of long wave-length radiations in identification work, it is pertinent to inquire how the relative intensities on long wave-length patterns compare with the intensities on patterns made with Mo$K\alpha$. Frevel has given an answer for Cu$K\alpha$ on a semi-empirical, semi-theoretical basis,[21] in the form of a table of conversion factors. This table shows that Cu$K\alpha$ intensities should not differ from Mo$K\alpha$ intensities by more than 25 percent, as long as the Cu$K\alpha$ Bragg angle is less than 45°. This difference is certainly not an important one. At Bragg angles greater than 45°, Cu$K\alpha$ relative intensities are very much higher than the relative intensities of lines of the same *d* value on a Mo$K\alpha$ pattern. There may be as much as a factor of 5 at high Bragg angle, arising from the Lorentz polarization factor. Because the most intense lines generally arise from the large *d* values which occur at small Bragg angle, it would seem that the wave-length effect on intensities could as a first approximation be neglected.

The above remarks and conclusion have necessarily been made on the basis of theory and apply to intensity *measurements*. Intensity estimates with Cu$K\alpha$ will of course be as good or as bad as Mo$K\alpha$ estimates. In addition there is the effect of absorption by the sample which can be

TABLE II. Comparison x-ray diffraction data taken from A.S.T.M. card file for anatase ($TiO_2$). The number of the card is given, II denoting a card from the First Supplement—1945. The third pair of columns is stated to represent data from two sources, in good agreement, with the symbol *n* denoting lines not given by both sources.

| Card 1406 Mo$K\alpha$-radiation |  | Card II-876 Mo$K\alpha$-radiation |  | Card II-911 Mo$K\alpha$-radiation |  |
|---|---|---|---|---|---|
| *d* | $I/I_1$ measured | *d* | $I/I_1$ estimated | *d* | $I/I_1$ estimated |
| 3.52 | 1.00 | 3.508 | 1.0 | 3.47 | 1.0 |
|  |  | 2.425 | .1 | 2.39 | .4 |
| 2.37 | .24 | 2.372 | .5 | 2.35 | .4*n* |
|  |  | 2.333 | .1 | 2.305 | .2*n* |
| 1.88 | .40 | 1.887 | .9 | 1.878 | .9 |
| 1.70 | .28 | 1.696 | .7 | 1.689 | .7 |
| 1.66 | .24 | 1.662 | .7 | 1.654 | .7 |
| 1.480 | .24 | 1.447 | .7 | 1.472 | .6 |
| 1.362 | .08 | 1.361 | .6 | 1.358 | .3 |
| 1.335 | .08 | 1.335 | .6 | 1.332 | .4 |
| 1.262 | .11 | 1.261 | .7 | 1.258 | .6 |
|  |  | 1.247 | .2 | 1.242 | .4*n* |
| 1.164 | .06 | 1.163 | .6 | 1.159 | .4*b* |
|  |  | 1.158 | .1 |  |  |
|  |  | 1.054 | .1 |  |  |
| 1.045 | .03 | 1.0488 | .5 | 1.044 | .3 |
|  |  | 1.0407 | .5 | 1.038 | .6*n* |
|  |  | 1.0155 | .5 | 1.014 | .2 |
|  |  | 1.0043 | .2 | 1.002 | .2 |
|  |  | .9939 | .1 |  |  |
| .950 | .02 | .9528 | .7 | .951 | .3 |
|  |  | .9437 | .6 | .943 | .4 |
| .913 | .02 | .9167 | .7 | .916 | .6 |
|  |  | .9114 | .7 | .912 | .3 |
| .894 | .02 |  |  | .892 | .1*n* |
|  |  |  |  | .876 | .1*n* |
|  |  |  |  | .843 | .1*n* |
|  |  |  |  | .824 | .1*n* |
|  |  |  |  | .806 | .1*n* |
|  |  |  |  | .795 | .1*n* |
|  |  |  |  | .740 | .1*n* |
|  |  |  |  | .702 | .1*n* |
|  |  |  |  | .668 | .1*n* |

---

[21] Quoted in the "Alphabetical Index of X-ray Diffraction Patterns," A.S.T.M., Philadelphia, Pennsylvania.

important with long wave-lengths. The effect of absorption is to reduce the intensity of high $d$ lines, and because these are generally the strong lines, to raise effectively all relative intensities. That this effect is present to a significant degree is entirely supported by eight cases of duplicate card file data involving both Mo$K\alpha$ and Cu$K\alpha$.

All of the above remarks for Cu$K\alpha$ intensity apply with much more force to the longer wave-lengths such as Fe$K\alpha$ and Cr$K\alpha$. With these radiations the absorption factor is exceedingly important, and further the important lines begin to occur at large Bragg angle where they are enhanced by the Lorentz polarization factor. Three cases have been found in the card file literature where Mo$K\alpha$ intensity measurements are duplicated by estimates on patterns of the same material made with Ni$K\alpha$, Fe$K\alpha$, and Cr$K\alpha$. The long wave-length intensities show only small differences from line to line, while the Mo$K\alpha$ intensities for corresponding lines show radical differences.

## VI. LIMITATIONS AND ADVANTAGES

The x-ray diffraction method of chemical identification is obviously limited to materials listed in the standards library available which ordinarily excludes solid solutions, and, of course, eliminates non-crystalling materials. In addition the method suffers the major limitation that it is not very sensitive to small *concentrations* of a given phase. The minimum concentration that will be detectable will depend upon the substance and also in part, upon the other substances present. Hanawalt states that the minimum may be as low as 1.0 percent or less, or more than 50.0 percent. Perhaps a representative figure would be around 10.0 percent. In general only experiments with known mixtures of the substances to be detected will definitely decide the minimum detectable concentration. However, it is clear that substances giving several lines of high absolute intensities will be detectable in smaller concentration than those yielding only weak lines.

Not uncommonly in industrial work a sample must be studied which contains four or five phases, some of which may be present in such small concentration as to show only one or two of their strongest lines. Clearly in cases of this kind a positive identification cannot be made by x-ray means alone. However, in practical cases some information is usually available to suggest some of the possible substances whose presence the x-ray pattern may confirm. Even in the most difficult cases the diffraction pattern usually yields some information which supplements that obtained from other methods and justifies the effort involved.

The x-ray method also fails to give complete identification in several minor instances. For example, there are numerous oxides of the type $AB_2O_4$ which have the same structure, the spinel type, and of these materials many have lattice parameters that are nearly the same. In addition then to nearly identical $d$ values, the intensities turn out to be similar. Clearly an identification beyond that of the structure type is difficult in such a case. The spinel situation emphasizes that the x-ray method identifies *structures*, from which rigid chemical analyses must not be inferred.

The x-ray diffraction method has unique advantages. Like the petrographic microscope it identifies the phases present and not merely the elements of which they are composed. Dimorphous forms of the same substance are as easily distinguished as different substances, an exceedingly important fact. Unlike the microscope it is not sensitive to very small concentrations of a minor phase but it is particularly successful with opaque materials and very finely divided samples.

Other major advantages of the x-ray method are that only a small sample is needed and the accuracy necessary is easy to achieve. The sample which is used is not destroyed so that it is available for chemical or spectroscopic study later. The x-ray method can upon occasion rival the companion electron diffraction method with respect to its ability to obtain results from a small amount of specimen.

## VII. AUXILIARY PROCEDURES

X-ray diffraction will not displace other means of examination but it does supplement them with information not otherwise obtainable and in some cases it may yield the same information with less effort. In difficult cases it is most helpful to have supplemental information obtained by spectroscopic or conventional chemical analysis. The major cation constitution is particularly useful in interpretation as well as being a time saver if the cations are such as to cause fluorescence. Industrial clients are occasionally reluctant to furnish such analytical and source information as they already possess regarding a sample, with the result that the x-ray worker is greatly handicapped in interpreting his results.

The writers prefer to give each sample a preliminary examination under the petrographic microscope. Sometimes this makes x-ray examination unnecessary. It is frequently possible to tell at a glance whether the sample is homogeneous or not and at times the presence of a minor phase may be revealed which would be overlooked by x-ray examination alone. Where several phases are present it may be desirable to make a preliminary separation by magnetic means, by heavy liquids, by washing or some other physical means and then examine the fractions separately by x-rays.

It has been stated on certain occasions that the x-ray diffraction method permits chemical identification work to be done with untrained personnel. Probably this is true where a certain number of routine determinations of the same type are to be done over and over again. However, it cannot be too strongly stressed that good x-ray diffraction work demands a high degree of experience and judgment, and at least a reasonable background in x-ray theory and crystallography, together with an understanding of the industrial problems to which the work is being applied. It will be completely futile for any laboratory to purchase an expensive x-ray diffraction outfit and expect to "hire a boy" to operate it.

# Charts for Computation of *d*-Values in X-Ray Diffraction Chemical Analysis*

Orley E. Brown
*Department of Mathematics, Case School of Applied Science, Cleveland 6, Ohio*

(Received October 16, 1946)

In chemical identification by x-ray diffraction, Bragg's equation

$$2d = \lambda \csc\theta,$$

in which $d$ represents the crystallographic interplanar spacing, $\lambda$ the x-ray wave-length, and $\theta$ the Bragg angle, must be repeatedly solved to find the $d$-value which corresponds to a known $\lambda$-value and an observed $\theta$-value. Whereas this equation lends itself to conventional treatment graphically, it is not convenient to secure the accuracy required by the ordinary graphical methods.

Since only five different values of $\lambda$, one for each of five different x-ray targets, are commonly used, these five values were assigned and a doubly graduated scale was prepared for each of the five resulting equations. This type of chart has the advantage of being constructable for any desired accuracy and of being immune to damage due to paper shrinkage or distortion in reproduction.

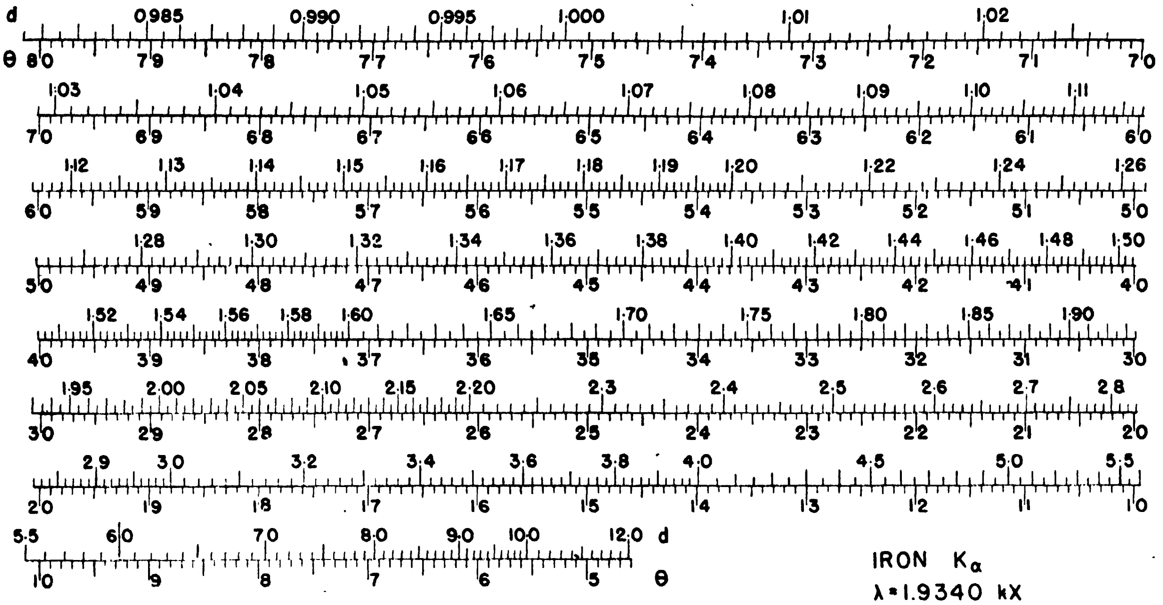
To use one of the charts, locate the $\theta$-value on the lower side of the scale and read off the corresponding $d$-value from the upper side.

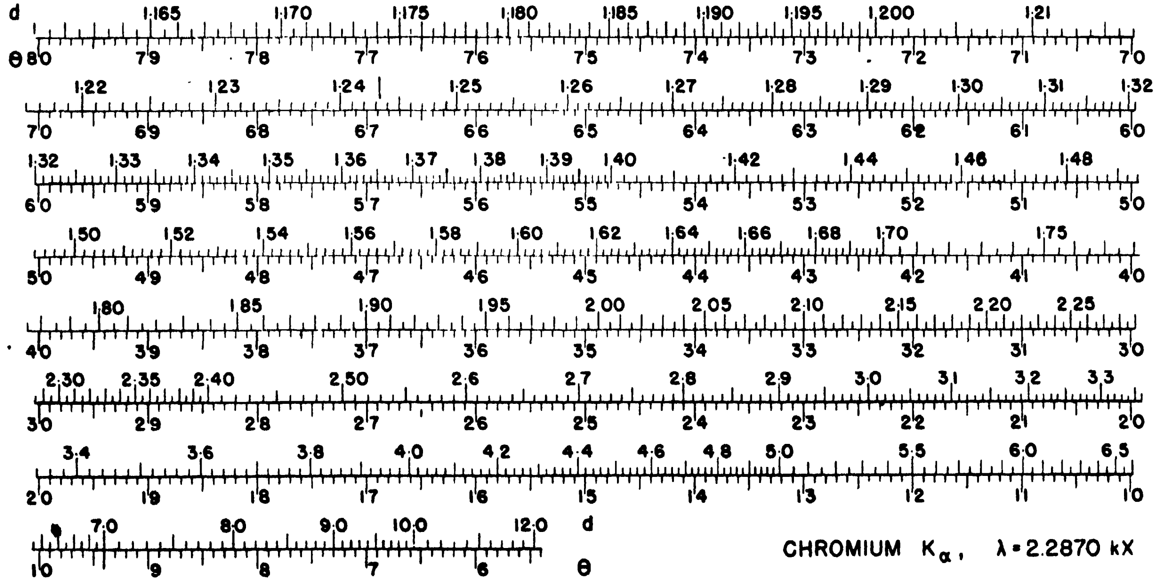
The plates used in printing these charts are being retained, and if the demand warrants, arrangements will be made with a commercial concern to supply reprints of the charts in quantity.

---

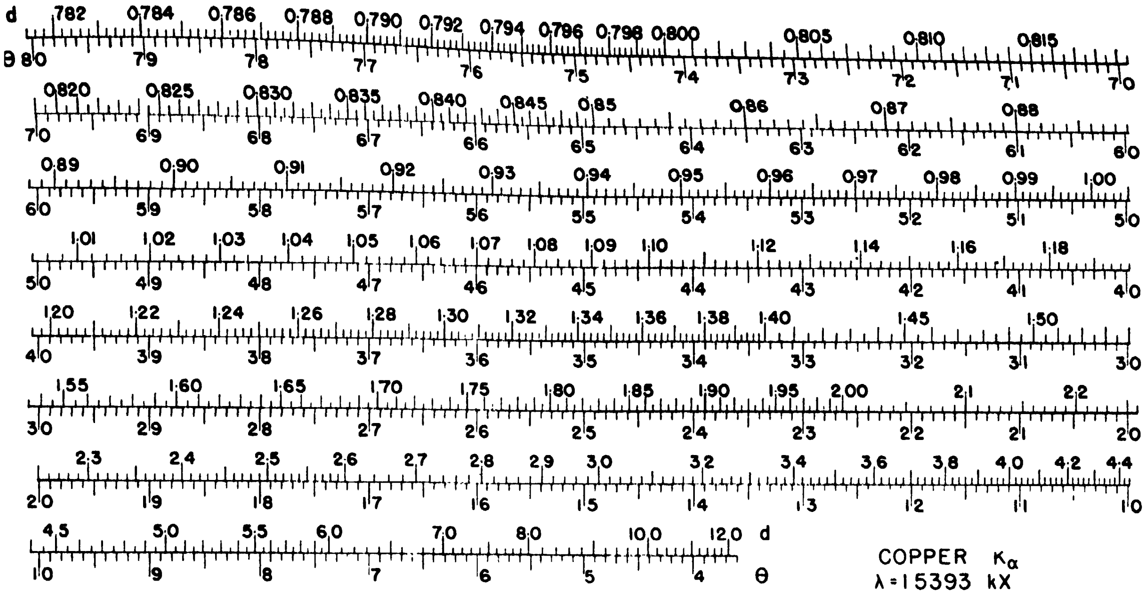
* These charts were undertaken at the suggestion of Richard L. Barrett. See paper in this issue by Smith and Barrett, p. 177.

## Charts for Computation of *d*-Values in X-Ray Diffraction Chemical Analysis

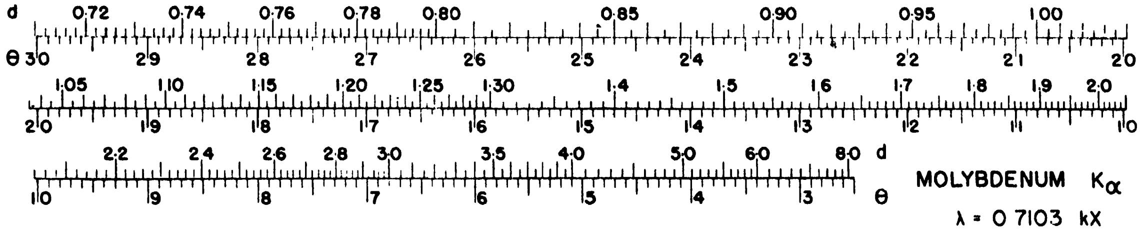




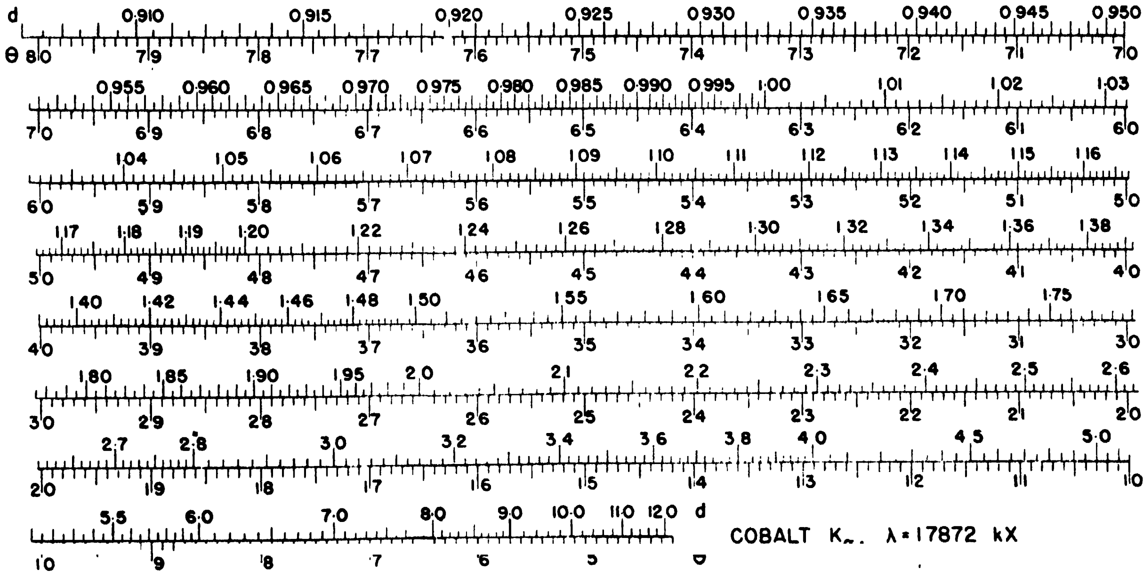
# Charts for Computation of *d*-Values in X-Ray Diffraction Chemical Analysis



COPPER $K_\alpha$
$\lambda = 1.5393$ kX



MOLYBDENUM $K_\alpha$
$\lambda = 0.7103$ kX



COBALT $K_\alpha$. $\lambda = 1.7872$ kX

# Streamlines for the Subsonic Flow of a Compressible Fluid Past a Sphere

E. R. VAN DRIEST*
*Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Received June 28, 1946)

The second approximation to the velocity potential for the irrotational flow of a compressible fluid past a sphere, derived independently by Tamada and Kaplan, is used to calculate the subsonic streamlines for the flow of air at a Mach number of 0.5 past a sphere. Comparison is made between these streamlines and those for the irrotational flow of an incompressible fluid. Velocity and pressure comparisons are also shown.

## INTRODUCTION

THE subsonic irrotational flow of a compressible fluid past a sphere has been studied extensively by Lord Rayleigh,[1] K. Tamada,[2] and C. Kaplan.[3] Designating the flow of an incompressible fluid past a sphere as the zero approximation to a solution expressed in ascending powers of $M^2$ for the flow of a compressible fluid about the same object, the first approximation was obtained by Lord Rayleigh and the second by Tamada and Kaplan. Inasmuch as the work of Tamada and Kaplan included the comparison of velocity and pressure distribution on the surface of the sphere for compressible fluid flow with the corresponding distributions for the incompressible flow, the author thought it would be also worth while to make a comparison of the streamlines for the two cases. The present paper therefore shows the second approximation to the streamlines for the flow of air at a Mach number of 0.5 and the corresponding streamlines when the fluid is considered incompressible.

## THE SECOND APPROXIMATION TO THE POTENTIAL FUNCTION

In the case of the steady adiabatic frictionless flow of a perfect gas, the local velocity of sound is given by

$$c^2=c_0^2\left[1+\frac{\gamma-1}{2}M^2\left(1-\frac{V^2}{U^2}\right)\right], \qquad (1)$$

where $U$ is the velocity of the undisturbed stream, $c_0$ the velocity of sound corresponding to the undisturbed stream, $M$ the Mach number for the undisturbed stream $(=U/c_0)$, $\gamma$ the ratio of the specific heats of the fluid, and $V$ the local velocity of the fluid.

Furthermore, assuming irrotational motion, combination of the equations of motion and continuity takes the form

$$\Delta\phi=\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}+\frac{\partial^2\phi}{\partial z^2}$$
$$=\frac{1}{2c^2}\left(\frac{\partial\phi}{\partial x}\frac{\partial V^2}{\partial x}+\frac{\partial\phi}{\partial y}\frac{\partial V^2}{\partial y}+\frac{\partial\phi}{\partial z}\frac{\partial V^2}{\partial z}\right), \qquad (2)$$

where $\phi$ is the velocity potential. When the motion has axial symmetry as for the flow past a sphere, Eq. (2) becomes in polar coordinates $(r, \theta)$

$$\Delta\phi=\frac{\partial^2\phi}{\partial r^2}+\frac{2}{r}\frac{\partial\phi}{\partial r}+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\phi}{\partial\theta}\right)$$
$$=\frac{1}{2c^2}\left(\frac{\partial\phi}{\partial r}\frac{\partial V^2}{\partial r}+\frac{1}{r^2}\frac{\partial\phi}{\partial\theta}\frac{\partial V^2}{\partial\theta}\right), \qquad (3)$$

in which

$$V^2=\left(\frac{\partial\phi}{\partial r}\right)^2+\frac{1}{r^2}\left(\frac{\partial\phi}{\partial\theta}\right)^2. \qquad (4)$$

By reason of the extreme complexity of Eq. (3), methods for approximate solution have been resorted to. As stated in the introduction, a first approximation was obtained by Lord Rayleigh; he took the well-known solution for incompressible flow as the zero approximation, substituted it into the right-hand side of Eq. (3), placed $c=c_0$, and then obtained a solution of the new equation.

---

* Assistant Professor of Mechanical Engineering.

[1] Lord Rayleigh, "On the flow of compressible fluid past an obstacle," Phil. Mag. 32, No. 187, 1–6 (July, 1916).

[2] K. Tamada, "On the flow of a compressible fluid past a sphere," Proc. Phys.-Math. Soc., Japan 21, 743–752 (1939).

[3] C. Kaplan, "Flow of a compressible fluid past a sphere," N. A. C. A. Tech. Note, No. 762 (May, 1940).

The second approximation to the compressible fluid flow, accomplished by Tamada and Kaplan, was carried out under the assumption that $\phi$ can be expanded as a power series in $M^2$, thus

$$\phi=\phi_0+\phi_1 M^2+\phi_2 M^4+\cdots. \tag{5}$$

Substituting Eq. (5) into Eq. (3) and equating terms of like powers of $M$ on both sides, there results

$$\Delta\phi_0=0, \tag{6}$$

$$\Delta\phi_1=\frac{1}{2U^2}\left(\frac{\partial\phi_0}{\partial r}\frac{\partial V_0^2}{\partial r}+\frac{1}{r^2}\frac{\partial\phi_0}{\partial\theta}\frac{\partial V_0^2}{\partial\theta}\right), \tag{7}$$

$$\begin{aligned}\Delta\phi_2={}&\tfrac{1}{2}(\gamma-1)(V_0^2-1)\Delta\phi_1\\&+\frac{1}{2U^2}\left(\frac{\partial\phi_0}{\partial r}\frac{\partial V_1^2}{\partial r}+\frac{1}{r^2}\frac{\partial\phi_0}{\partial\theta}\frac{\partial V_1^2}{\partial\theta}\right)\\&+\frac{1}{2U^2}\left(\frac{\partial\phi_1}{\partial r}\frac{\partial V_0^2}{\partial r}+\frac{1}{r^2}\frac{\partial\phi_1}{\partial\theta}\frac{\partial V_0^2}{\partial\theta}\right).\end{aligned} \tag{8}$$

It will be noted that the solution of the first equation represents the incompressible fluid case, *viz.*,

$$\phi_0=U\left(r+\frac{a^3}{2r^2}\right)P_1, \tag{9}$$

wherein $a$ is the radius of the sphere and $P_1$ is the Legendre polynomial of the first order (of argument $\cos\theta$). The second equation leads to the first-order correction to the velocity potential as obtained by Lord Rayleigh, i.e.,

$$\begin{aligned}\phi_1=a^3U\Bigg[&\left(\frac{1}{3}\frac{1}{r^2}-\frac{1}{5}\frac{a^3}{r^5}+\frac{1}{24}\frac{a^6}{r^8}\right)P_1\\&+\left(-\frac{3}{10}\frac{1}{r^2}+\frac{27}{55}\frac{a^2}{r^4}-\frac{3}{10}\frac{a^3}{r^5}+\frac{3}{176}\frac{a^6}{r^8}\right)P_3\Bigg].\end{aligned} \tag{10}$$

The third equation yields the second-order correction as given by Tamada and Kaplan:

$$\begin{aligned}\phi_2=a^3U(\gamma-1)\Bigg[&\left(\frac{599}{16800}\frac{1}{r^2}-\frac{3}{70}\frac{a^3}{r^5}+\frac{1}{24}\frac{a^6}{r^8}-\frac{13}{560}\frac{a^9}{r^{11}}+\frac{13}{2800}\frac{a^{12}}{r^{14}}\right)P_1+\left(\frac{83133}{3665200}\frac{a^2}{r^4}-\frac{1}{10}\frac{a^3}{r^5}+\frac{15}{176}\frac{a^6}{r^8}\right.\\&\left.-\frac{57}{1960}\frac{a^9}{r^{11}}+\frac{23}{6800}\frac{a^{12}}{r^{14}}\right)P_3+\left(\frac{1}{7}\frac{a^3}{r^5}-\frac{50517}{276640}\frac{a^4}{r^6}+\frac{3}{52}\frac{a^6}{r^8}-\frac{9}{1120}\frac{a^9}{r^{11}}+\frac{5}{8512}\frac{a^{12}}{r^{14}}\right)P_5\Bigg]\\&+a^3U\left(\frac{1325952}{4065600}\frac{1}{r^2}-\frac{49}{150}\frac{a^3}{r^5}-\frac{243}{1925}\frac{a^5}{r^7}+\frac{1049}{3080}\frac{a^6}{r^8}+\frac{5589}{84700}\frac{a^8}{r^{10}}-\frac{3049}{18480}\frac{a^9}{r^{11}}+\frac{7789}{369600}\frac{a^{12}}{r^{14}}\right)P_1\\&+a^3U\left(-\frac{53}{150}\frac{1}{r^2}+\frac{70620063}{95295200}\frac{a^2}{r^4}-\frac{61}{150}\frac{a^3}{r^5}-\frac{156}{275}\frac{a^5}{r^7}+\frac{28}{55}\frac{a^6}{r^8}+\frac{1137}{7150}\frac{a^8}{r^{10}}-\frac{12631}{64680}\frac{a^9}{r^{11}}+\frac{929}{56100}\frac{a^{12}}{r^{14}}\right)P_3\\&+a^3U\left(\frac{5}{42}\frac{1}{r^2}-\frac{6}{11}\frac{a^2}{r^4}+\frac{2}{3}\frac{a^3}{r^5}+\frac{135921}{640640}\frac{a^4}{r^6}-\frac{60}{77}\frac{a^5}{r^7}+\frac{253}{728}\frac{a^6}{r^8}+\frac{87}{1540}\frac{a^8}{r^{10}}-\frac{997}{18480}\frac{a^9}{r^{11}}+\frac{85}{29568}\frac{a^{12}}{r^{14}}\right)P_5.\end{aligned} \tag{11}$$

## DETERMINATION OF THE STREAMLINES

The potential and stream functions are connected by the relations

$$\frac{1}{r}\frac{\partial\phi}{\partial\theta}=-\frac{1}{\rho r\sin\theta}\frac{\partial\psi}{\partial r}, \tag{12}$$

$$\frac{\partial\phi}{\partial r}=\frac{1}{\rho r^2\sin\theta}\frac{\partial\psi}{\partial\theta}, \tag{13}$$

in which

$$\rho=\rho_0\left[1+\frac{\gamma-1}{2}M^2\left(1-\frac{V^2}{U^2}\right)\right]^{1/(\gamma-1)} \tag{14}$$

where $\rho_0$ is the density of the undisturbed stream.

From Eqs. (12), (14), and (4), there results

$$\rho_0\left\{1+\frac{\gamma-1}{2}M^2\left[1-\frac{1}{U^2}\left(\frac{\partial\phi}{\partial r}\right)^2-\frac{1}{U^2r^2}\left(\frac{\partial\phi}{\partial\theta}\right)^2\right]\right\}^{1/(\gamma-1)}\sin\theta\frac{\partial\phi}{\partial\theta}=-\frac{\partial\psi}{\partial r}, \tag{15}$$

or, absorbing $\rho_0$ in $\psi$,

$$\left\{1+\frac{\gamma-1}{2}M^2\left[1-\frac{1}{U^2}\left(\frac{\partial\phi}{\partial r}\right)^2-\frac{1}{U^2r^2}\left(\frac{\partial\phi}{\partial\theta}\right)^2\right]\right\}^{1/(\gamma-1)}\sin\theta\frac{\partial\phi}{\partial\theta}+=-\frac{\partial\psi}{\partial r}. \quad (16)$$

For the particular case of $M=0.5$ and $\gamma=1.405$ for air, Eq. (16) becomes

$$\left\{1+0.0506\left[1-\frac{1}{U^2}\left(\frac{\partial\phi}{\partial r}\right)^2-\frac{1}{U^2r^2}\left(\frac{\partial\phi}{\partial\theta}\right)^2\right]\right\}^{2.465}\sin\theta\frac{\partial\phi}{\partial\theta}=-\frac{\partial\psi}{\partial r}. \quad (17)$$

Now the simplest way of integrating Eq. (17) is by graphical means, i.e., for certain values of $\theta$, the left-hand side can be computed at various values of $r$, and the area under the curve can be calculated to give corresponding values of $\psi$. The graph relating $\psi$ and $r$ can then be used to determine the radial distance $r$ for certain preassigned values of $\psi$. This process can of course be carried out to any degree of accuracy.

Upon differentiation of Eq. (5) and substitution of $M=0.5$,

$$\frac{\partial\phi}{\partial\theta}=\frac{\partial\phi_0}{\partial\theta}+\frac{1}{4}\frac{\partial\phi_1}{\partial\theta}+\frac{1}{16}\frac{\partial\phi_2}{\partial\theta}. \quad (18)$$

Substitution of Eqs. (9)–(11) into Eq. (18) then yields, when $a=1$,

$$\begin{aligned}\frac{\partial\phi}{\partial\theta}=&-U\sin\theta(r+0.5699r^{-2}+0.0556r^{-4}\\&-0.0995r^{-5}+0.0020r^{-6}-0.0326r^{-7}\\&+0.0525r^{-8}+0.0087r^{-10}-0.0166r^{-11}\\&+0.0019r^{-14})\\&-U\sin3\theta(-0.1760r^{-2}+0.2899r^{-4}\\&-0.1560r^{-5}+0.0071r^{-6}-0.1064r^{-7}\\&+0.0909r^{-8}+0.0215r^{-10}-0.0272r^{-11}\\&+0.0023r^{-14})\\&-U\sin5\theta(0.0183r^{-2}-0.0840r^{-4}\\&+0.1116r^{-5}+0.0212r^{-6}-0.1200r^{-7}\\&+0.0570r^{-8}+0.0087r^{-10}-0.0088r^{-11}\\&+0.0005r^{-14}).\end{aligned} \quad (19)$$

Furthermore,

$$\frac{\partial\phi}{\partial r}=\frac{\partial\phi_0}{\partial r}+\frac{1}{4}\frac{\partial\phi_1}{\partial r}+\frac{1}{16}\frac{\partial\phi_2}{\partial r},$$

so that

$$\begin{aligned}\frac{\partial\phi}{\partial r}=&\,U\cos\theta(1-1.1394r^{-3}-0.2224r^{-5}\\&+0.4975r^{-6}-0.0121r^{-7}+0.2280r^{-8}\\&-0.4201r^{-9}-0.0866r^{-11}+0.1822r^{-12}\\&-0.0266r^{-15})\\&+U\cos3\theta(0.1181r^{-3}-0.3870r^{-5}\\&+0.2595r^{-6}-0.0142r^{-7}+0.2481r^{-8}\\&-0.2416r^{-9}-0.0716r^{-11}+0.0996r^{-12}\\&-0.0106r^{-15})\\&+U\cos5\theta(-0.0073r^{-3}+0.0671r^{-5}\\&-0.1116r^{-6}-0.0255r^{-7}+0.1679r^{-8}\\&-0.0915r^{-9}-0.0174r^{-11}+0.0194r^{-12}\\&-0.0013r^{-15}).\end{aligned} \quad (20)$$

This is all the further that the process can be reduced. The remaining work is pure computation. In this paper, streamlines were located to the nearest one-thousandth of the unit for each of the following nine angles of $\theta$: $\tan^{-1} 1/15$, $\tan^{-1} 1/8$, $\tan^{-1} 1/4$, $\tan^{-1} 4/10$, $\pi/6$, $\pi/4$, $\pi/3$, $\tan^{-1} 11/3$, and $\pi/2$.

Figure 1 shows the streamlines for both the compressible and incompressible fluid flow. Also in this figure are plotted the corresponding velocities in the field at $\theta=\pi/2$ and the corresponding pressures on the surface of the sphere. (Note that in the figure $u_0=U$, and $u=V$ in the field at $\theta=\pi/2$.) The pressure distribution for the compressible fluid case is given by the formula

$$\frac{p-p_0}{\frac{1}{2}\rho_0U^2}=\frac{1}{(\gamma/2)M^2}\times\left\{\left[1+\frac{\gamma-1}{2}M^2\left(1-\frac{V^2}{U^2}\right)\right]^{\gamma/(\gamma-1)}-1\right\}, \quad (21)$$

where $p_0$ is the pressure in the undisturbed

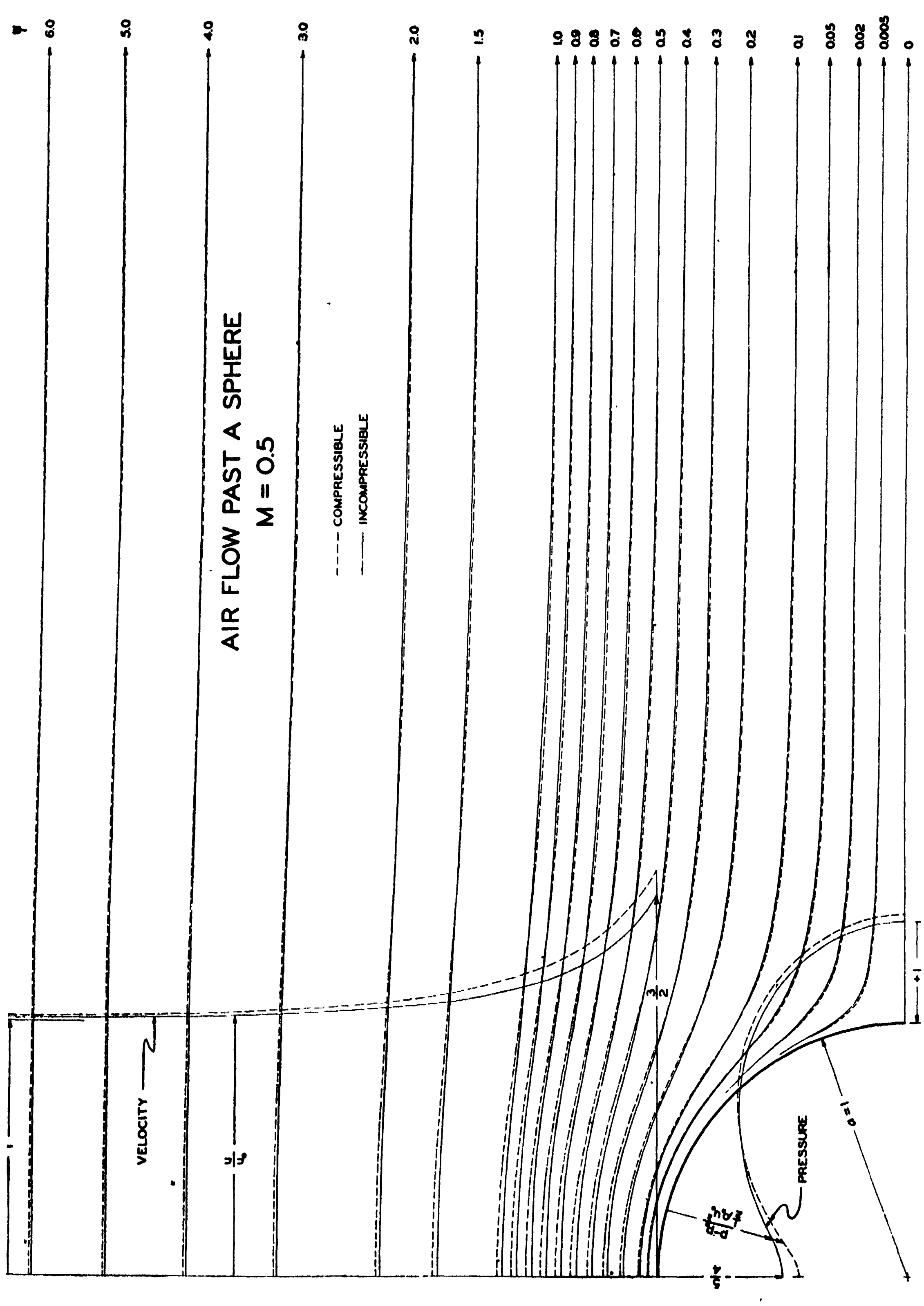


FIG. 1. Streamlines for the flow of both a compressible and an incompressible fluid past a sphere.

TABLE I.

| | on surface of sphere | | | | in field at $\theta=\pi/2$ | | |
|---|---|---|---|---|---|---|---|
| | $\frac{p-p_0}{\frac{1}{2}\rho_0 U^2}$ | | $\frac{V}{U}$ | | | $\frac{V}{U}$ | |
| $\theta$ | comp. | incomp. | comp. | incomp. | $r$ | comp. | incomp. |
| 0 | +1.070 | +1.000 | 0 | 0 | 1.0 | 1.600 | 1.500 |
| $\tan^{-1} 1/15$ | +1.058 | +0.990 | 0.093 | 0.100 | 1.1 | 1.464 | 1.376 |
| $\tan^{-1} 1/8$ | +1.030 | +0.965 | 0.174 | 0.186 | 1.2 | 1.368 | 1.289 |
| $\tan^{-1} 1/4$ | +0.930 | +0.867 | 0.342 | 0.364 | 1.4 | 1.241 | 1.182 |
| $\tan^{-1} 4/10$ | +0.751 | +0.689 | 0.528 | 0.557 | 1.8 | 1.120 | 1.086 |
| $\pi/6$ | +0.496 | +0.437 | 0.719 | 0.750 | 2.2 | 1.069 | 1.047 |
| $\pi/4$ | −0.098 | −0.125 | 1.049 | 1.061 | 3.0 | 1.029 | 1.019 |
| $\pi/3$ | −0.736 | −0.688 | 1.331 | 1.299 | 4.0 | 1.012 | 1.008 |
| $\tan^{-1} 11/3$ | −1.218 | −1.093 | 1.526 | 1.447 | | | |
| $\pi/2$ | −1.416 | −1.250 | 1.600 | 1.500 | | | |

stream, or

$$\frac{p-p_0}{\frac{1}{2}\rho_0 U^2}=5.695\left\{\left[1+0.0506\left(1-\frac{V^2}{U^2}\right)\right]^{3.465}-1\right\} \quad (22)$$

for the air flow at Mach number 0.5.

Table I gives the velocity and pressure data for the two flows.

It is seen that in the case of the compressible fluid the point on the sphere at which the pressure is equal to the pressure in the undisturbed stream is at a location different from that for the incompressible fluid. For the compressible fluid flow the point is located by placing $-U=(1/r)(\partial\phi/\partial\theta)$, substituting Eq. (19) for $\partial\phi/\partial\theta$, putting $r=1$, and then solving the resulting equation, *viz.*,

$$1.5419\sin\theta-0.0539\sin 3\theta+0.0045\sin 5\theta=1. \quad (23)$$

The solution of Eq. (23) is $\theta=43°\,38'$. For the incompressible fluid case, the point of zero pressure difference is located at $\theta=41°\,49'$.

The local Mach number is greatest at the top of the sphere. This number can be computed using Eq. (1), thus

$$M_l=\frac{V}{c}=\frac{V}{c_0\left[1+\frac{\gamma-1}{2}M^2\left(1-\frac{V^2}{U^2}\right)\right]^{\frac{1}{2}}}, \quad (24)$$

or, since $c_0=U/M$,

$$M_l=\frac{V/U}{\left[\frac{1}{M^2}+\frac{\gamma-1}{2}\left(1-\frac{V^2}{U^2}\right)\right]^{\frac{1}{2}}}. \quad (25)$$

At the top of the sphere where $V/U=1.600$, $M_l=0.833$, and therefore there is no shock in the field of flow. This is to be expected since the critical Mach number of the undisturbed flow for shock establishment is approximately 0.55 as indicated in another paper by Tamada.[4]

---

[4] K. Tamada, "Further studies on the flow of a compressible fluid past a sphere," Proc. Phys.-Math. Soc., Japan, **22**, 519–525 (1940).

# Minimum Sparking Potentials of Barium, Magnesium, and Aluminum in Argon

HAROLD JACOBS AND ARMAND P. LAROCQUE
*Sylvania Electric Products Inc., Kew Gardens, New York, New York**

(Received August 21, 1946)

Minimum sparking potentials of barium, magnesium, and aluminum in argon are found to be 93±½ v, 123±1 v, and 155±1 v, respectively. Utilizing the data in measuring the minimum sparking potentials, the ratios of electrons emitted from the cathode per bombarding positive ion ($\gamma_m$) are determined and found to be .149±.001 for barium, .089±.033 for magnesium, and .045±.002 for aluminum. A relationship is found such that, at a given field per unit pressure, a lower work function indicates a lower sparking potential and a higher $\gamma_m$ indicates a lower sparking potential, for the case of the three metals studied. A graphical method is proposed for predicting the minimum sparking potentials for metals, whose sparking potential values have not been determined, based upon the values of $\gamma_m$ or the work function $\phi$.

WHEN considering the design and manufacture of electron tubes which contain gases at reduced pressure, it has been found that the surface phenomenon plays a major role in successful operations and applications. For instance, the sparking potential of many gas tubes has been found to be a function of the cathode surface as well as the gas in the region of the minimum sparking potential. Similarly, the voltage drop across the gas tube while in operation is a function of the surface as well as the gas.

In spite of the large number of applications of gas tubes, relatively little is known concerning the actual mechanism of operation of these devices. This is probably due to the large number of variables which are possible in gas tubes, and to the fact that techniques of measurement have not been adequately developed. For instance, it has been found that small traces of mercury,[1] oil vapors, and oxygen in the gas or adsorbed by the surfaces in a tube will cause great alterations in the behavior of the tube. Thus, reproduceable results have been difficult to obtain, and, actually, it has only been in recent years that any valid data has been obtained.

One of the problems concerning gas tubes, then, is the effect of surfaces in gases upon sparking potential. In the following experiments, the surfaces of Ba, Mg, and Al were studied in reduced pressures of argon and the minimum sparking potentials obtained. From this data, various interpretations have been made including the deduction of the ratio of emitted electrons released by bombarding positive ions and the importance of this ratio upon the sparking potential.

## EXPERIMENTAL METHOD

Tubes were constructed as sketched in Fig. 1. It should be noted that these tubes consisted of two electrodes and a getter container which held the metal to be studied. Each of the two electrodes were made of nickel and shaped in the

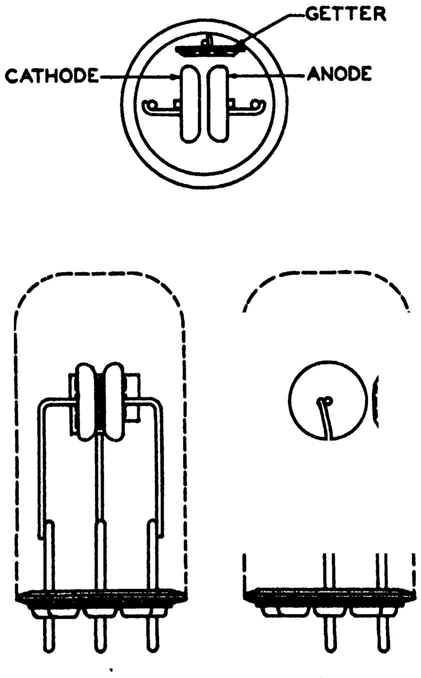


FIG. 1. Experimental tube.

---

* This work was undertaken in association with the Applied Physics Laboratory of the Johns Hopkins University under Section T series contract NOrd 7872 sponsored by the Bureau of Ordnance, United States Navy.

[1] W. E. Bowls, Phys. Rev. 53, 293 (1938).

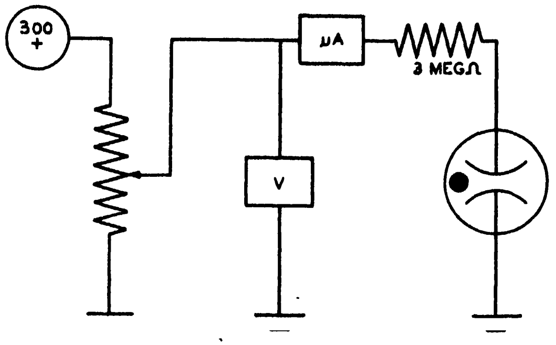


FIG. 2. Circuit for measuring sparking potentials.

manner of "Rogowsky electrodes."[2] That is, the electrodes were shaped such that surfaces facing each other followed equipotential lines. This was done to prevent concentrations of field at the edges and would make quantitative measurements of the field readily accessible. The tube to be studied was placed on exhaust together with a "dummy tube" containing only a barium-aluminum getter. This second tube was used as a reservoir to down any impurities in the argon, i.e., $O_2$, $N_2$, $H_2$ which was to be later admitted.

The nickel electrodes and getter container were then degassed as well as possible and the metal to be studied was evaporated on the nickel surface. In the case of barium, an iron clad barium getter was used and was electrically flashed. Magnesium powder was used for the Mg surface and aluminum powder for the Al surface. These reactions in no way depend upon reducing agents and hence the chance for impurity was greatly reduced. Argon was then admitted to the system from a standard Airco argon bottle (99.9+ percent pure). Liquid air was used at all times between the mercury manometer and the vacuum system. The electrical circuit used was such as sketched in Fig. 2. The criteria for sparking was such that a sudden rise in current would indicate the "breakdown" through the tube. In all cases, tests were conducted for stability of the surface before measurements were made. This was done by allowing 3μa current to pass through the tube. If current and voltage conditions were constant with respect to time, the surface was assumed to be stable. No measurement was made of the thickness of the metal deposited on the nickel cathode.

## DATA

The curves in Fig. 3 indicate the results of measuring sparking potential in relation to pressure times distance. In the case of Ba, Mg, and Al in argon, it was found that the minimum sparking potentials were $93\pm\frac{1}{2}$, $123\pm1$, and $155\pm1$ volts, respectively.

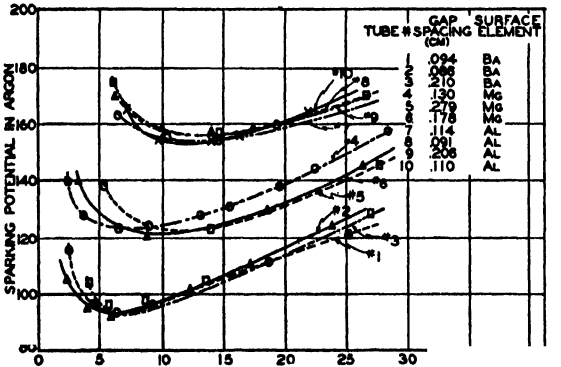
## PREVIOUS DATA CONCERNING COLD CATHODES

Before attempting to interpret the above data, it would be of interest to consider the work of previous experimenters in the field of cold cathodes.

It has been demonstrated,[3] that in tubes such as studied in these experiments, the following relation exists

$$i=i_0\frac{e^{\alpha x}}{1-\gamma(e^{\alpha x}-1)}. \tag{1}$$

$i$ = total current,
$i_0$ = constant,[4]
$x$ = distance between electrodes,
$\alpha$ = ratio of ions formed to electrons in a unit distance,
$\gamma$ = roughly, ratio of secondary emission from the cathode to the ions bombarding the cathode.



| TUBE # | GAP SPACING (CM) | SURFACE ELEMENT |
|---|---|---|
| 1 | .094 | BA |
| 2 | .086 | BA |
| 3 | .210 | BA |
| 4 | .130 | MG |
| 5 | .279 | MG |
| 6 | .178 | MG |
| 7 | .114 | AL |
| 8 | .091 | AL |
| 9 | .206 | AL |
| 10 | .110 | AL |

FIG. 3. Sparking potentials of barium, magnesium, and aluminum in argon *vs.* pressure×distance. The ordinate identifications should read 0, .5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0.

---

[2] J. D. Cobine, *Gaseous Conductors* (McGraw-Hill Book Company, Inc., New York, 1941), 171–181.

[3] L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases* (J. Wiley & Sons, Inc., New York, 1939), 378.

[4] The symbol $i_0$ which is the initial photoelectric current leaving the cathode is not truly constant and for very large values of $i$, there is found to be a slight drop in the

Now, this equation has been used by some experimenters as a criteria for determining or predicting sparking potentials. That is, if the denominator goes to zero, the current goes to infinity. Hence, a criteria for sparking may be:

$$1-\gamma(e^{\alpha x}-1)=0,\ \gamma=\frac{1}{e^{\alpha x}-1}. \tag{2}$$

As we shall see later, the higher the value of $\gamma$, the lower is the sparking potential in a given gas. Thus, the constant is extremely important from the point of view of any quantitative work on low pressure cold cathode discharge tubes. The exact interpretation of $\gamma$ has not been well developed, but it should be emphasized that although $\gamma$ is roughly the ratio of emitted electrons per positive ion bombarding the cathode, it also contains the ratio of electrons per quanta of light producing photoelectrons, and possibly several other sources of electrons which have been assumed to be small in comparison with the first ratio. Because of the importance of $\gamma$ in gas discharge work, there has been some research in recent years to determine the nature of $\gamma$ and the mechanism involved.

As late as 1938, Bowls[5] opened many questions concerning the constant $\gamma$. He utilized a mercury free vacuum system, platinum and sodium cathodes, and nitrogen gas. He obtained his values of $\gamma$ not by sparking but by measuring current and computing $\gamma$ from the following equation.

$$i=i_0\frac{(\alpha-\beta)e^{(\alpha-\beta)x}}{\alpha-\beta e^{(\alpha-\beta)x}}, \tag{3}$$

where

$$\beta/\alpha=\gamma.$$

Three very important discoveries were brought to light.

1. There was an "anomalous" peaking of $\gamma$ at low $X/p$ values. This had never been observed by sparking potential measurements.

2. When Hg was introduced to the system, radically different results occurred as to the value of $\gamma$. Thus, it appears that all work done previously without safeguards as to mercury contamination may have been in serious error.

3. In nitrogen, the Na cathode had a lower $\gamma$ than the Pt cathode. This is contrary to what would be expected since the lower work function cathode should have the higher $\gamma$ and hence the lower sparking potential.

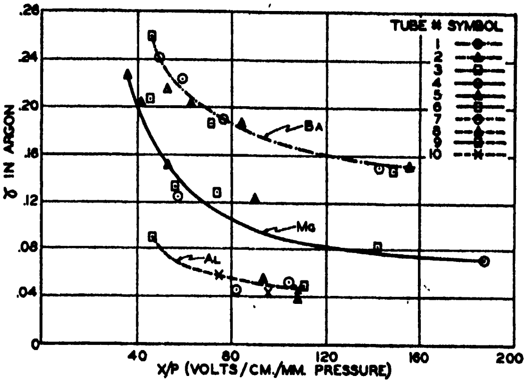


FIG. 4. $\gamma$ of barium, magnesium, and aluminum in argon *vs.* field per unit pressure.

In 1939, Hale[6] continued this work, but this time made measurements in hydrogen instead of nitrogen. His results again indicated:

1. An "anomalous" peaking in $\gamma$ appeared at very low $X/p$ values. (Pt: $X/p=125$; NaH: $X/p=10$.)

2. The presence of Hg could alter the characteristics of the surface completely.

3. Except at very low $X/p$ ratios, the $\gamma$ of the sodium was less than that for platinum.

Now, in this work, it was thought that the anomalous peaking was caused by a photoelectric phenomenon. That mercury vapor could change the surface was well established. The question of work function *versus* $\gamma$ was by no means settled and still remains open to question. In the experiments of Bowls and Hale, it is possible that the nitrogen and hydrogen combined with the Na surface in such a manner as to create a high work function surface. The answers to this question will be revealed by further experiments in inert gases.

In 1939, Huxford[7] studied Cs-Ag-O surfaces in photo-tubes filled with argon. He measured $\gamma$ by two methods, using the sparking criteria method as indicated above and also the method of measuring current prior to discharge and calculating the $\gamma$-values by means of Eq. (1). He obtained very good agreement using both methods.

Another piece of work is described by Loeb[8]

---

sparking potential. However, for very small currents ($<1\ \mu$a) such as used in the experiments above, the effect is small and has not been detected experimentally. For further comment, see L. B. Loeb, reference 3, pp. 420–425.

[5] W. E. Bowls, reference 1, 293.

[6] D. H. Hale, Phys. Rev. 55, 815 (1939).

[7] W. S. Huxford, Phys. Rev. 55, 754 (1939).

[8] L. B. Loeb, reference 3, p. 53.

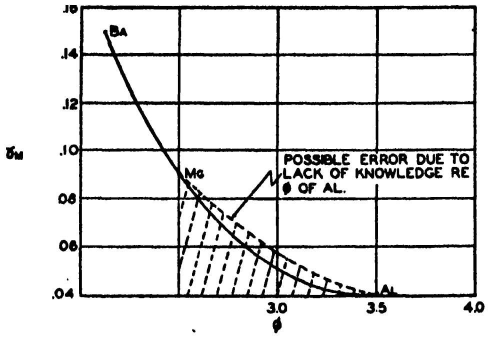


FIG. 5. $\gamma_m$ (ratio of emitted electrons per bombarding positive ion at the minimum sparking potential) *vs.* work function.

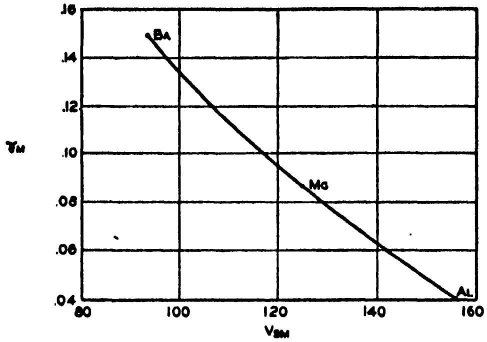


FIG. 6. $\gamma_m$ (ratio of emitted electrons per bombarding positive ion at the minimum sparking potential) as a function of the minimum sparking potential.

concerning some work done by Engstrom[9] on barium cathodes in argon. Here, also, good agreement is obtained between the two methods of measuring $\gamma$ except that the sparking method provides a smoother line. In other words, it is claimed that when measuring by sparking methods, the anomalies are wiped out and a more continuous value of $\gamma$ is obtained as a function of $X/p$.

## INTERPRETATION OF DATA

With the above background in mind, we can now return to the data concerning Ba, Mg, and Al in argon. Utilizing Eq. (2) again:

$$\gamma=\frac{1}{e^{\alpha x}-1},\quad X=\frac{V_s}{x},$$
$$\therefore\quad \gamma=\frac{1}{e^{\alpha}V_s/X-1},\quad \gamma=\frac{1}{e^{(\alpha/p)\cdot(p/X)V_s}-1}. \tag{4}$$

$V_s$ = sparking potential,
$X$ = field (volts/cm),
$x$ = distance (cm),
$p$ = pressure (mm).

The values for $\gamma$ as determined by Eq. (4) are shown in Fig. 4. The $\gamma_m$ values (at the minimum sparking potential) were found to be .149±.001 for barium, .089±.033 for magnesium, and .045±.002 for aluminum.

Having evaluated $\gamma$, we are in a position to draw certain conclusions. The work functions of Ba,[10] Mg,[11] and Al[11] are 2.1 electron volts, 2.42 electron volts, and 3.57(2.5−3.6) electron volts, respectively.

We can therefore conclude from the data in Fig. 4, that, at any of the given values of $X/p$ in the range we have observed, and in the case of the metallic surfaces studied, a lower work function indicates a higher $\gamma$. And, in addition, from the data in Fig. 3, a lower work function will indicate a lower minimum sparking potential.

One may wonder, then, why Bowls obtained a lower $\gamma$ in a large region of $X/p$ for platinum than was obtained for sodium. The answer may well be that the Na was not free on his cathode surface, but was, in fact, combined to form a composite cathode with nitrogen.

We should emphasize that in the case of Ba, Mg, and Al we were dealing with pure metal surfaces. As yet, there are no available data concerning the effects of $\gamma$ and $\phi$ on composite cathodes in gas tubes.

Knowing $\gamma$ and the $\phi$ of the surfaces in argon, it may be possible to predict the $V_{sm}$ (minimum sparking potential) of a cold cathode gas tube. If these predictions are valid, the designer would take a tube with a given sparking requirement and estimate the region of sparking potential by considering the published values of the work functions of the surfaces. Loeb[12] gives the following formula for minimum sparking potentials.

$$V_{sm}=\frac{3B_4}{A_4}\log\frac{1}{\gamma_m}+\frac{B_4}{A_4}\,\frac{1}{\log\left(\frac{1}{\gamma_m}\right)}. \tag{5}$$

---

[9] L. B. Loeb, reference 3, p. 407; R. W. Engstrom, Phys. Rev. 56 (1939).

[10] A. L. Reimann, *Thermionic Emission* (J. Wiley & Sons, Inc., New York, 1934), p. 80–1.

[11] J. D. Cobine, reference 2, p. 109.

[12] L. B. Loeb, reference 3, p. 414.

$B_4$ and $A_4$ are constants which could be evaluated. However, at best, this is an empirical formula and does satisfy a requirement which can be fulfilled more rapidly and simply by graphical methods.

In the first graph, we have $\gamma_m$ *versus* $\phi$. (Figure 5.)

Then by either knowing $\gamma_m$ or $\phi$, we may estimate $V_{sm}$ of a surface in argon by Fig. 6.

As an example of the use of these charts, suppose we calculate the $V_{sm}$ of sodium. The $\phi$ of sodium is given as 2.46−1.90 volts.[13] This is roughly a $\gamma_m$ in argon of from .08 to slightly more than .15 using Fig. 5. Now, turning to Fig. 6, this is roughly between 90 volts and 125 volts. According to Loeb,[14] the $V_{sm}$ of a Na surface in argon is 95 volts, or in the expected region.

We realize our range is limited and more samples may be desirable in future work, but the indications are that this may be a fruitful method for utilizing data concerning cold cathodes.

Acknowledgment should be made to Mr. Gerald Rich, Sylvania Electric Products, Inc., Kew Gardens, New York, for his discussions and help relating to this work.

---

[13] J. D. Cobine, reference 2, p. 109.

[14] L. B. Loeb, reference 3, p. 414.

---

# Fatigue of Ag-Cs$_2$O, Ag-Cs Photoelectric Surfaces

S. Pakswer
*Continental Electric Company, Geneva, Illinois*

(Received August 8, 1946)

Fatigue of Ag-Cs$_2$O, Ag-Cs surfaces has been observed in blue and red light. In blue light the long wave threshold shifts to shorter wave-length and the long wave maximum increases and sometimes shifts to shorter wave-length. Rise of temperature at the cathode counteracts the phenomena. In red light temperature effects are dominating. The observed phenomena are in line with de Boer's theory of the photoelectric effect on such surfaces as the formation of positive ions on the surface causes polarization of the neighboring Cs-atoms and a change in their selective absorption.

SURFACES with *S*1* spectral sensitivity have been commonly manufactured for a number of years and the technique of their production is well known. There is, however, little agreement concerning the mechanism of the photoelectric emission. The spectral sensitivity characteristic of the *S*1 surface shows two main maxima—one at approximately 3500A and the other around 8000A corresponding probably to two different emission processes. It has not yet been definitely established where these electron emission processes take place.

According to the ideas developed by J. H. deBoer,[1,2] the sensitivity maximum at 8000A is caused by cesium atoms adsorbed at so-called "active" spots on the surface of a semi-conductive layer containing Cs$_2$O and interspersed with Cs and Ag metal atoms. The surface layer of the semi-conductor polarizes the atoms of Cs and thus reduces their work function. The emitted electrons are supplied by diffusion through the layer of the semi-conductor.

Timofeev,[3] in a series of experimental investigations, emphasizes the importance of the formation of positive charges on the surface of the semi-conductive layer, and their influence on the photoelectric and secondary electron emission of Ag-Cs cathodes.

Other authors[4-6] believe that the photo-ionization takes place inside the layer of the semi-conductor as an "inner" photoelectric effect and that the electrons have to break through the

---

* *S*1 and *S*4 spectral sensitivity are RMA designations for "red" and "blue" sensitive tubes, respectively.

[1] J. H. deBoer, *Electron Emission and Adsorption Phenomena* (Cambridge University Press, England, 1935).

[2] J. H. deBoer and M. C. Teves, Zeits. f. Physik **83**, 521 (1933).

[3] Timofeev, J. Tech. Phys. U.S.S.R. **6**, 340 (1944).

[4] H. Teichmann, Zeits. f. Electrochemie **44**, 602 (1938).

[5] G. Maurer, Zeits. f. Physik **60**, 170 (1941).

[6] N. S. Chlebnikov, Bull. Acad. Sci. U.S.S.R. Ser. Phys. **8**, 286 (1944).

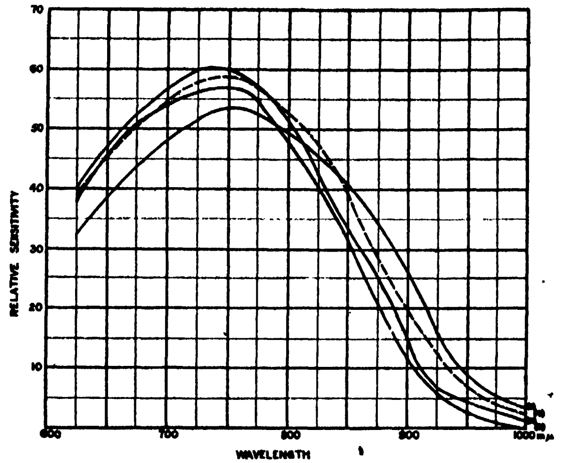


FIG. 1. Tube No. 10 type CE-1 vacuum exposed to blue light (200 w at 1′+filter 554) curve 1: 0 hr., curve 2: 1 hr., curve 3: 48 hr. illumination, curve 4 (———): 78 hr. in the dark.

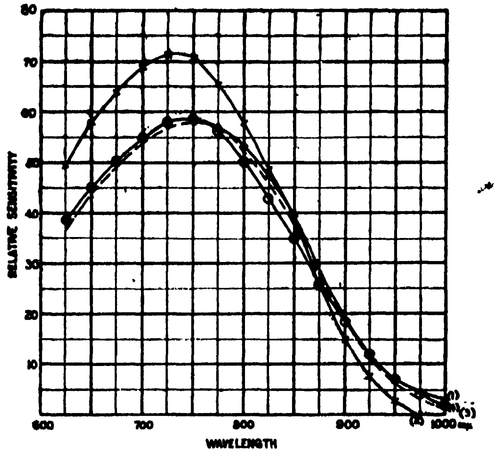


FIG. 2. Tube No. 10 type CE-1 vacuum exposed to blue light (500 w at 12″+filter 554) curve 1 (0): 0 hr., curve 2 ($X$): 1 hr., curve 3 (△): 24 hr. illumination, curve 4 (———): 50 hr. in the dark.

potential barrier at the surface, possibly at "active" spots.[5]

Experiments tend to support the deBoer conception[7] particularly the behavior in polarized light, the high velocity of emitted photoelectrons and the low quantum yield. The semi-conductor theory is supported by experience won on the secondary emission, by the spectral characteristic of the inner photo-effect[4] and by the influence of the thickness of the oxide layer on the sensitivity and the spectral characteristic of $S1$ surfaces.[8] Electron microscopic study of $S1$ surfaces,[9] showing spotty, non-uniform distribution of electrons emitted under illumination by light of 6000A, can be explained by both theories.

An insight into the mechanism of photoelectric emission can be obtained by a study of the fatigue of $S1$ surfaces. DeBoer[10] observed that tubes irradiated with blue light, without voltage applied, dropped in sensitivity to white incandescent light more than tubes irradiated with green light, and that red light and infra-red rays not only did not cause fatigue, but regenerated the original sensitivity of tubes which lost sensitivity. With voltage applied and a white light source, the drop of sensitivity was a function of the applied voltage and the illumination on the cathode. The fatigue phenomena increased at low temperature (−196°C) and decreased at higher temperature. In tubes exposed to white light of .15 lumen, the infra-red threshold was shifted considerably to shorter wave-lengths and the response in the maximum was lower and the maximum itself was shifted slightly to shorter wave-lengths.

To explain these phenomena deBoer assumed that the positive cesium ions recombine with electrons diffusing to them from the conductivity levels of the semi-conductor. The rate of recovery from fatigue is given by the rate of diffusion. Illumination with infra-red rays or rise of temperature increase the thermal agitation of

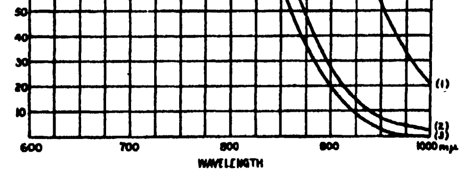


FIG. 3. Tube No. 4 CE-1 gas-filled exposed to blue light (200-w lamp at 1′+filter 554) $V_A = 22.5$ v curve 1: 0 hr., curve 2: 3 hr., curve 3: 264 hr. illumination.

---

[7] G. Lewin, Trans. Electrochem. Soc. 87, (1945), preprint.

[8] P. M. Morosov and M. M. Butslov, Bull. Acad. Sci. U.S.S.R., Ser. Phys. 8, 291 (1944).

[9] Yosimaro Morija, Electrotech. J. (Japan) 1, 65 (1937).

[10] J. H. deBoer and M. C. Teves, Zeits. f. Physik 74, 604 (1932).

the lattice, and thus the amount of electrons diffusing toward the surface. If voltage is applied, positive cesium ions migrate toward the cathode into the layer and must be replaced by cesium atoms migrating from the layer to the surface. A decrease of free alkali atoms at the surface has also been observed by Timofeev and Kondorskaya[11] who found that the time variation in fatigue phenomena is a series of exponentials.

It was felt that the data given in literature do not give a complete picture of fatigue phenomena. Following observation of the influence of different parameters on the stability and the fatigue of Ag-Cs photo-tubes, it was, therefore, decided to obtain additional data. As sensitivity measurements in white light give only integrated values, it seemed that a study of spectral distribution would give a better picture of the phenomena.

Measurements of the distribution of spectral sensitivity were made with a Coleman spectrophotometer Model 10S, calibrated against a VanCittert monochromator and a thermopile. The instrument was a Leeds & Northrup galvanometer. The plate voltage was 22.5 v. The color temperature of the lamp in the photometer was 2870°K. All tests were made on tubes of the type CE-1/918, vacuum and gas-filled, with silver-plated copper cathode and also on gas-filled 918 tubes with massive silver cathode.

### TESTS IN "BLUE" LIGHT

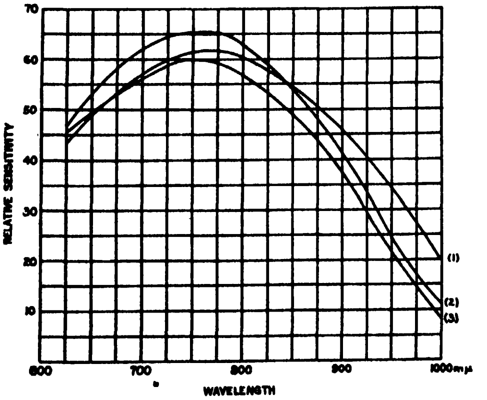


FIG. 4. Tube No. 3a. 918 gas-filled exposed to blue light (200-w lamp at 12″ filter 554) $V_A = 22.5$ v curve 1: 0 hr., curve 2: 3 hr., curve 3: 264 hr. illumination.

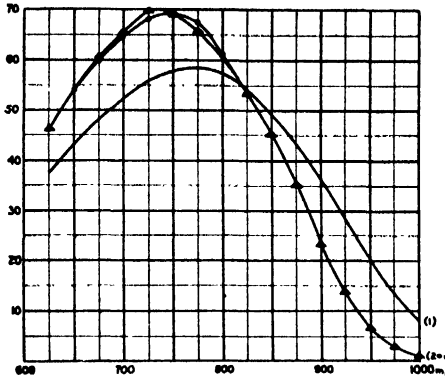


FIG. 5. Tube No. 16 CE-1 vacuum, curve 1: 0 hr., curve 2: exposed 1 hr. to blue light (500-w lamp at 1′+filter 554), curve 3: 2 hr. in dark at −80°C.

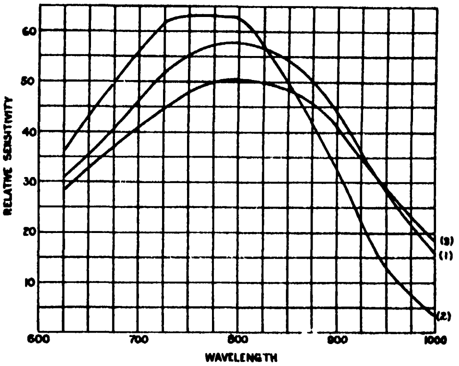


FIG. 6. Tube No. 17 CE-1 vacuum, curve 1:0 hr., curve 2: 1 hr. exposed to blue light (500-w lamp at 1′+filter 554), curve 3: 2 hr. in dark at +90°C.

The tubes were exposed to the light of a 200-w lamp at 12″ distance filtered by a Corning filter 554; only blue and ultraviolet radiation fell on the cathode. In some tests the 200-w lamp was replaced by a 500-w projection lamp. The intensity of the radiation in this case, as measured with a blue-sensitive CE-61 tube was increased three times.

In Figs. 1 and 2 are represented typical distribution curves. In Fig. 1 curve 1 is the original curve. After 1-hour illumination by the 200-w lamp without any plate current on, the sensitivity in the maximum surprisingly increased considerably and the threshold shifted to shorter wave-lengths (curve 2). After a protracted period of time the sensitivity at the maximum decreased slightly (curve 3, after 48 hr.) and the threshold was shifted still further. The tube was now kept in the dark. After 78 hours (curve 4) the increase at the maximum still persisted

---

[11] Timofeev and Kondorskaya, Physik. Zeits. Sovjetunion 9, 683 (1936).

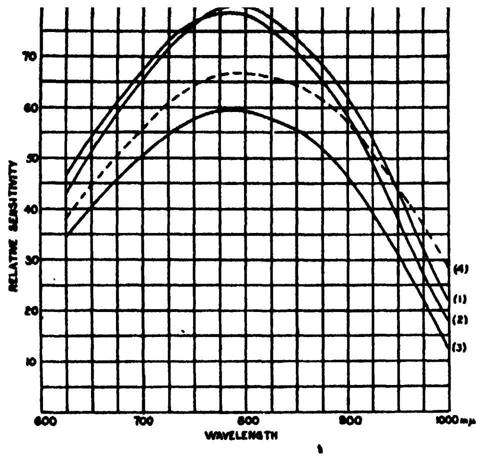


FIG. 7. Tube No. 15 type CE-1 vacuum exposed to red light (500-w lamp at 1′+filter 244) curve 1: 0 hr., curve 2: 1 hr., curve 3: 24 hr., curve 4: 50 hr. in the dark.

but the threshold moved back to longer wave-lengths.

This tube was now subjected to the light of the 500-w lamp (Fig. 2). After 1 hour the maximum had greatly increased (curve 2) but fell down rapidly (curve 3); the tube was heated up to about 50°C. After 50 hours in the dark (curve 4) the maximum and the threshold shifted to longer wave-lengths.

All other tubes investigated, gas-filled and vacuum, silver and silver plated, showed this initial increase in the maximum. Sometimes a shift in the maximum could be observed (Fig. 3). In Fig. 4 is shown the response of a tube with a silver cathode.

It was thought that these tests indicated the influence of three separate processes:

1. A process in blue light leading to a shift and an increase of the maximum and a shift of the threshold.
2. A process due to the rise of temperature which decreases the maximum.
3. A process in the dark which shifts the threshold back to longer wave-lengths.

To investigate further, separately, the influence of light and temperature two tubes were exposed to light of the 500-w lamp for 1 hour (curve 2, Figs. 5 and 6). Then one tube was kept for two hours in a mixture of solid $CO_2$ and acetone ($-80°C$) (curve 3, Fig. 5) and the other at $+90°C$ for the same period of time (curve 3, Fig. 6). In the first tube the initial changes persisted while in the second tube the sensitivity in the maximum decreased considerably and the sensitivity on the long wave side recuperated completely.

### TESTS IN "RED" LIGHT

The same lamps were used in conjunction with Corning filter 244 transmitting red-infra-red radiation. With the 200-w lamp the curves showed a small drop of sensitivity in all wave-lengths. With the 500-w lamp (Fig. 7) where the heat development on the cathode was considerable, the sensitivity in the maximum went down considerably after 24 hours (curve 3). A large amount of sensitivity, especially at long wave-lengths, was recuperated after 50 hours in the dark.

### CONCLUSION

It can be easily understood that changes of spectral sensitivity, combining with the energy distribution of an incandescent lamp, will cause variations of output current which will appear as "instability." Parameters affecting the long wave end of the sensitivity curve will have a relatively greater effect than parameters affecting the maximum. This will explain why "blue" light or high plate voltage cause a higher drop of sensitivity than red light. By proper choice of the color temperature of the lamp, the temperature, the applied voltage, and the intensity of the light, it is possible to balance the different influences out and obtain stable operation.

It is difficult to explain the observed phenomena solely by the recombination of positive ions and electrons. It can be easily shown that fatigue and recovery would follow exponential or hyperbolic laws in case of recombination and depending on its mechanism.

The increase and the shift of the maximum and the shift of the threshold indicate changes in the electrical state of the Cs-atoms on the surface due to the formation of positive charges. This polarization of atoms next to an ion, and the change of selective absorption of light following the Franck-Condon principle, have been described by deBoer in his book,[1] mainly on pages 193, 228, 280, and 333.

Acknowledgment is made to Mr. John A. Loeber, who performed most of the experiments described above.

# The Measurement of Elasticity in Fluids of Low Viscosity

John R. Van Wazer* and Herbert Goldberg**
*Department of Manufacturing Experiments, Eastman Kodak Company, Rochester, New York*

(Received August 19, 1946)

By means of an apparatus applying an alternating shearing motion to a liquid, it was shown that fluids with a viscosity of around one poise have elastic properties. A discussion of the variation of the resonance maxima with the amount of damping is given, and the shear modulus and damping viscosity are calculated for some dilute solutions of high polymeric materials.

## INTRODUCTION

Although all forms of matter exhibit elastic behavior upon the application of hydrostatic force, it is customary to think of only solids and rather thick gels as being elastic on the application of force in shear or extension. This viewpoint is based on the concept that the relaxation time of the structural bonds holding a liquid together is of the order of $10^{-8}$ sec. However, a different state of affairs exists in many fluids. When a close-knit gel of a high polymer is diluted, the interlocking molecular chains to which the elasticity in shear is attributed are gradually separated and broken. It thus appears that the shear modulus of such a gel will decrease upon dilution until there are not enough continuous polymer chains reaching from wall to wall to give a measurable elastic effect. This type of elasticity can be observed experimentally when the time in which inter-polymer bonds make or break is greater than or of the same order of magnitude as that of the experiment—that is to say, when the relaxation time is measurably long.

Kendall[1] has shown that some high polymeric solutions can be diluted to very low concentrations and still exhibit elastic recoil in shear. Although Kendall's method is very sensitive, it suffers from the difficulties of interpretation inherent in transient measurements. In the study reported below, the oscillatory method of Goldberg and Sandvik[2] has been adapted to the measurement of shear elasticity in fluids of low viscosity. This latter method appears to be as sensitive as that of Kendall and is, of course, more easily interpreted.

## EXPERIMENTAL PROCEDURE

The apparatus used in these experiments consisted primarily of a pair of concentric cylinders between which was placed the liquid under study. The outer cylinder was made from a piece of glass tubing selected for regularity of bore. It had a diameter of 1.240 cm. This cylinder was held in the apparatus so that it could be made to oscillate about its axis at a given amplitude and any chosen frequency. The frequency was adjusted by changing the speed of a variable speed motor which drove the oscillating cylinder through a cam and rocker-arm device. Photographic study showed that the rotatory oscillations of the outer cylinder were pure sine waves. The frequency of rotation of the outer cylinder was measured by a strobotac lamp.

The hollow inner cylinder (4.55 cm long with a diameter of 1.174 cm), having a low moment of inertia (4.12 g-cm$^2$) about its major axis, was suspended from a thin fiber. This low moment of inertia was achieved by building the cylinder of thin-walled tubing affixed to a heavy central shaft which served to weight the cylinder so that it would not float. A phonograph needle was used as a bottom bearing for the inner cylinder. Before an experiment was started, the cylinders were aligned by adjusting the support holding the suspension thread; then the inner cylinder was lifted slightly away from its bearing. A mirror was attached to the central rod of the inner cylinder so that a beam of light could be used to amplify its motion. The suspension fiber consisted of a single strand of untwisted silk approximately 15 cm long. Thus, the elasticity of the fiber was so slight that its contributions to the

---

* Now in the Research Laboratory, Rumford Chemical Works, Rumford, Rhode Island.

** Member of the Research Department, Eastman Kodak Company.

[1] Rheol. Bull. 12, 26 (1941).

[2] To be published.

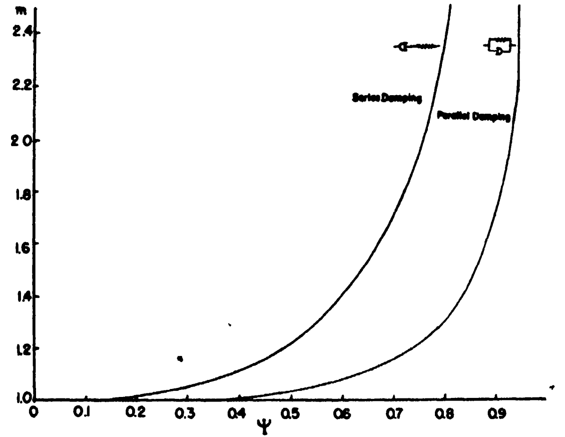


FIG. 1. Calculated variation of $\psi$ with $m$ at resonance for an elastic element with series or parallel damping.

observed elastic effects were negligible. In the studies reported below, end effects were negligible since the layer of liquid at either end of the inner cylinder was approximately 20 times thicker than that in the annular space between the two cylinders.

To make a measurement, the outer cylinder was filled with the solution under study, and then the inner cylinder was carefully inserted so that no air bubbles were entrapped between the two cylinders. After the cylinders were aligned, the outer cylinder was set in oscillation, and the amplitude of rotation of the inner cylinder was measured by the path of a beam of light reflected from its mirror onto a screen which was usually placed several meters from the instrument. Thus, any slight drifting of the inner cylinder with respect to the outer one was greatly amplified with the result that a slow erratic drifting was superimposed upon the oscillatory motion of the light spot. This necessitated "on the fly" measurement of the amplitude of oscillation with a caliper. When a non-elastic liquid was placed between the two cylinders, the inner cylinder had the same amplitude of rotation as that of the outer cylinder at low frequencies; and there was a gradual decrease in amplitude as the frequency was raised until at extremely high frequencies the inner cylinder remained practically motionless. When an elastic liquid was being studied, the amplitude of rotation of the inner cylinder was the same as that of the outer cylinder at very low frequencies and became negligibly small at high frequencies. However, at an intermediate frequency the amplitude of rotation of the inner cylinder was greater than that of the outer cylinder because of the phenomenon of resonance. It should be noted that the existence of mechanical resonance means that there is an elastic restoring force resisting distortion of the liquid. Therefore, the studies reported in this paper cannot be interpreted on the basis of thixotropy, simple non-Newtonian flow, or any other of the multitudinous rheological classifications which are concerned with changes in the internal friction of a substance.

The amplitude of rotation of the outer cylinder could be measured in two ways: first, by extrapolating the plot of amplitude *vs.* frequency of the inner cylinder to zero frequency, and secondly, by rigidly joining the two cylinders together and observing the light beam. The same value for the amplitude was always obtained by both of these methods.

The modulus of rigidity can be calculated from the following easily derived equation:

$$G_0=\frac{2\pi I}{r_1 L}\left(\frac{1}{r_1}-\frac{1}{r_2}\right)f_{\max}{}^2 \qquad (1)$$

where $G_0$ is the modulus of rigidity assuming no damping, $f_{\max}$ the resonance frequency, $L$ the length of the inner cylinder, and $r_1$ and $r_2$ are the radii of the inner and outer cylinders respectively.

The magnification factor, $m$, can be defined as the ratio of the amplitude of oscillation of the inner cylinder to that of the outer cylinder.

TABLE I. A comparison of elastic data from recoil and resonance studies.

| Material | Approximate viscosity (poises) | Kendall's resilient effect | Resonance data | | |
|---|---|---|---|---|---|
| | | | Shear modulus (dynes/cm²) | Type of damping | Damping viscosity (poises) |
| Lucite in dichlorobenzene | 1 | Very large | 80 | Series | 1 |
| Rubber cement in benzene | 1 | Large | 0.28 | Parallel | 0.028 |
| Boiled linseed oil | | None | No resonance maximum | | |
| Glucose in water | 1 | None | No resonance maximum | | |

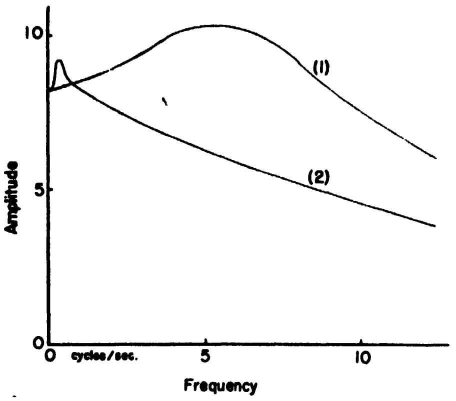


FIG. 2. Resonance curves: (1) methyl methacrylate polymer in dichlorobenzene; (2) rubber cement in benzene.

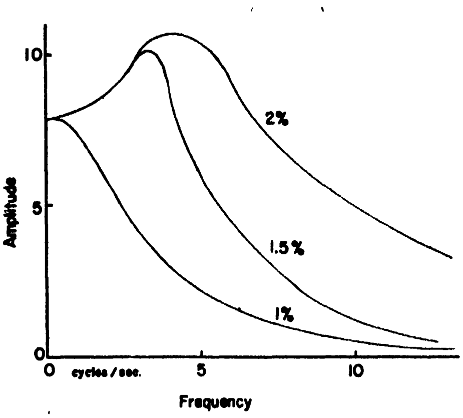


FIG. 3. Resonance curves for various concentrations of napalm in gasoline.

When the magnification factor is smaller than two or three, the modulus of rigidity computed from the previous equation is too small, because the damping causes a pronounced shift in the resonance frequency, and the equation is derived for an undamped system. It is impossible to compute the shift of the resonance maximum due to damping, unless a model for the system is assumed. We have calculated this shift for two models: a spring with parallel damping, and a spring with series damping. The calculations which are based on the direct electrical analog[3] are given in the appendix, and a graph of the ratio of the resonance frequency with damping to the resonance frequency that would be found, if the spring were not damped, *vs.* the magnification factor, is to be found in Fig. 1. From this graph the shear modulus as determined from the above equation can be corrected as follows:

$$G = G_0/\psi^2 \qquad (2)$$

where $G$ is the real shear modulus (assuming that the model is correct); $G_0$ is the shear modulus calculated from Eq. (1); and $\psi$ is the ratio of the frequencies taken from the graph. It is possible to compute the variation of $m$ with $\psi$ for the more complex model of a spring with a series and parallel dashpot, but instead of one $m-\psi$ curve, a family of curves will result depending on the relative values of the damping constants. However, this procedure was not followed because the apparent increase in accuracy of interpretation should not be accepted without fitting the whole frequency-amplitude curve.

The damping viscosity can be computed for the assumed models since the ratio of the damping viscosity to the shear modulus is related to the resonance frequency and magnification factor by means of Eqs. (2), (3), (6), and (9) in the Appendix, for the spring with either series or parallel damping. If the damping viscosity as calculated for series damping is much less than the viscosity determined in a conventional viscosimeter, the predominant damping is in parallel with the elastic element. If the calculated series viscosity is equal to the regular viscosity, the damping is in series with the elastic element. The calculated series viscosity cannot be greater than the regular viscosity. Thus, by comparison of the viscosity as calculated from the models with the regular viscosity, it is possible to find the predominant type of damping and thus the most correct value of the shear modulus as computed for a given fluid. In this discussion of damping, anomalous viscosity has been neglected since most fluids of low viscosity are only slightly non-Newtonian.

## DATA AND DISCUSSION

Several substances studied by Kendall were used with this apparatus to obtain the resonance curves shown in Fig. 2. The rheological constants derived from these measurements are presented in Table I.

An interesting series of measurements was performed on dilute sols of an aluminum soap

---

[3] J. Miles, J. Acous. Soc. Am. 14, 183 (1943).

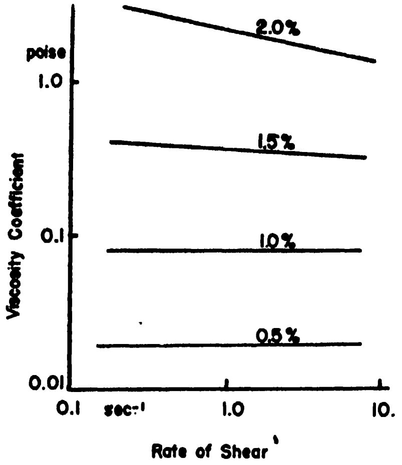


FIG. 4. The viscosity of napalm solutions at various concentrations.

in gasoline. The resonance curves for several concentrations are shown in Fig. 3. The constants calculated from these curves are presented in Table II. The predominant damping at a concentration of 1.5 percent is caused by the series viscosity. Measurements by Goldberg[4] indicate that the damping at concentrations higher than 6 percent is exclusively caused by the parallel viscosity. This means that both the viscosities in series and in parallel with the elastic element grow smaller with decreasing concentration, as might be expected. The transition from parallel to series damping of the elasticity occurs at about 2 percent. It is interesting to note that the concentration at which there is no longer any perceptible change of apparent viscosity coefficient with rate of shear (see Fig. 4) is also the concentration at which the resonance effect disappears.

The variation of shear modulus with molecular weight for a series of fractionated methyl methacrylate polymers is also given in Table II. Although the exact values of the maxima were difficult to determine because of the flatness of the resonance curves, it appears that the shear modulus increases with decreasing molecular weight at concentrations of equal viscosity. This result is in accord with the concept that the elastic properties of such a dilute solution are caused by the stretching of molecules.

TABLE II. Collected elastic data on fluids of low viscosity.

| Fluid studied | | Experimental data | | Calculated constants | | | | | Real constants | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Shear modulus (dynes/cm²) | | | Damping viscosity (poises) | | | | |
| Solute | Solvent | Magnification factor | Resonance frequency (cycles/sec.) | No damping | Series damping | Parallel damping | Series | Parallel | Viscosity of fluid (poises) | Predominant type of damping | Approximate shear modulus (dynes/cm²) |
| 1% napalm (aluminum soap) | gasoline | 1.01 (if any) | 0.6 (?) | 0.4 | 1.6 | 2.6 | 0.46 | 2.3 | 0.02 | Impossible | Not elastic |
| 1.5% napalm | gasoline | 1.29 | 3.5 | 11 | 36 | 18 | 0.76 | 0.88 | 0.8 | Series | 36 |
| 2.0% napalm | gasoline | 1.37 | 4.5 | 18 | 51 | 26 | 0.95 | 0.91 | 2± | Series and parallel | ca. 40 |
| 1% methyl methacrylate (mol. wt. = $32\times10^4$) | ortho dichlorobenzene | 1.31 | 5.3 | 25 | 81 | 38 | 1.2 | 1.25 | 2.0± | Series | 80 |
| 1.05% methyl methacrylate (mol. wt. = $32\times10^4$) | dichlorodiethyl sulfide (Levinstein HS) | 1.29 | 7.5 | 48 | 160 | 77 | 1.6 | 2.3 | 2.5± | Series | 160 |
| 1.65% methyl methacrylate (mol. wt. = $16\times10^4$) | same | 1.09 | 6.0 | 30 | 210 | 75 | 1.6 | 3.5 | 2.0 | Series | 210 |
| 1.95% methyl methacrylate (mol. wt. = $12.4\times10^4$) | same | 1.05 | 5.5 | 21 | 210 | 71 | 1.6 | 5.5 | 2.0 | Series | 210 |
| 2.85% methyl methacrylate (mol. wt. = $8.4\times10^4$) | same | 1.02 | 5.0 | 21 | 530 | 100 | 3.8 | 9.6 | 3.0 | Series | 530 |
| 4.76% methyl methacrylate (mol. wt. = $4.0\times10^4$) | same | 1.00 | none | — | — | — | — | — | 3.0 | | Not elastic |
| Rubber cement | benzene | 1.15 | 0.40 | 0.14 | 0.70 | 0.28 | 0.01 | 0.028 | 1.0 | Parallel | 0.28 |

---

[4] To be published.

## ACKNOWLEDGMENT

The work reported above was completed in August, 1943 under the sponsorship of N.D.R.C. We wish to thank Dr. E. K. Carver for his helpful advice and suggestions about this problem.

## APPENDIX

## Derivation of the Correction to Shear Modulus Due to Damping

The following discussion is based on the direct electrical analogy, the fundamental definitions of which are listed below:

| *Mechanical Quantity* | *Electrical Analogy* |
|---|---|
| $x$ displacement | $q$ charge |
| $F$ force | $v$ potential |
| $1/G$ reciprocal of shear modulus | $C$ capacitance |
| $n$ viscosity | $R$ resistance |
| $M$ mass | $L$ inductance |
| $v$ velocity | $i$ current |

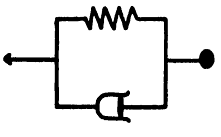
Because of the equivalence of Hooke's law to the capacity equation, the law of Newtonian viscosity to Ohm's law, and Newton's third law of motion to the equation for an inductance, this direct electrical analogy holds true.

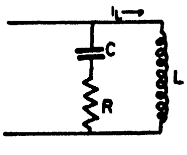
Since sine wave oscillations are applied to the apparatus, the displacement $x$ is related to the amplitude $X$ according to the following equation in which $t$ is the time, $\delta$ is a constant to take care of phase difference, and $\omega=2\pi f$ ($f=$frequency):

$$x_{\text{input}}=X_{\text{input}}\sin\omega t,$$
$$x_{\text{output}}=X_{\text{output}}\sin\omega(t+\delta).$$

By differentiation of the above equations with respect to time, we find that the ratio of the output to the input velocity equals the ratio of the output to the input amplitudes. This ratio is called the magnification factor and is denoted by the symbol $m$.

## Parallel Damping


Mechanical model



Electrical model

For this model undergoing sine wave oscillations we can set up the following equation for the magnification factor according to the direct electrical analog:

$$m=\frac{i_L}{i}=\frac{R-j/\omega C}{R+j(\omega L-1/\omega C)}. \tag{1}$$

This can be reduced to a more general form by use of $\psi=$the ratio of the observed damped frequency to the free natural frequency of the elastic element, and $d$, a damping coefficient.

$$d=R(C/L)^{\frac{1}{2}} \tag{2}$$

and since the free natural frequency equals $1/(LC)^{\frac{1}{2}}$

$$\psi=\omega(LC)^{\frac{1}{2}}. \tag{3}$$

Then

$$m=\frac{d-j/\psi}{d+j(\psi-1/\psi)}. \tag{4}$$

or

$$|m|^2=\frac{d^2+1/\psi^2}{d^2+\psi^2+1/\psi^2-2}. \tag{5}$$

If we differentiate and set $\partial m/\partial\psi$ equal to zero, the condition for a maximum, or resonance, is fulfilled and we find that

$$d^2=2(1-\psi^2)/\psi^4 \text{ at resonance.} \tag{6}$$

Equation (5) gives us the relation between $\psi$ and the physically measurable $m$. If we compute the shear modulus as if there were no damping, we obtain a quantity, $G_0$, according to the equation

$$G_0=k(f')^2.$$

But we really wish to know the true shear modulus

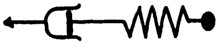
$$G=kf^2$$

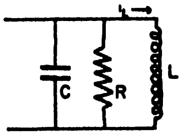
where $f'$ is the observed frequency with damping, and $f$ is the free natural frequency.

By definition $\psi=f'/f$:

$\therefore$ $G=G_0/\psi^2$. (This is Eq. (2) in the text.)

## Series Damping





The equation for the magnification factor with the substitutions of $\psi$ and $d$, defined in Eqs. (2) and (3) above, is given below.

$$m=\frac{i_L}{i}=\frac{-j/\psi}{1/d+j(\psi-1/\psi)}, \tag{7}$$

and

$$|m|^2=\frac{(1/\psi)^2}{1/d^2+\psi^2+1/\psi^2-2}. \tag{8}$$

Differentiating with respect to $\psi$ and setting $\partial m/\partial\psi=0$, we find that

$$d^2=1/2(1-\psi^2) \text{ at resonance.} \tag{9}$$

From Eq. (9)

$$m^2=1/[\psi^2(3\psi-4)+1] \text{ at resonance.} \tag{10}$$

This relationship can be used to correct $G_0$ to $G$ as was shown above. Equations (5) and (10) are graphed in Fig. 1.

# On Creep and Relaxation*

B. Gross
*National Institute of Technology, Rio de Janeiro, Brazil*

(Received June 17, 1946)

The creep and the relaxation function of linear systems, for which the principle of superposition is valid, are mutually connected in a simple way. This makes it possible to calculate the distribution function of relaxation times of stress and the distribution function of retardation times of strain, when the relaxation function or the creep function is given. It also allows a transformation formula to be established for the conversion of one distribution function into another. The results of the theory are applied to a detailed discussion of the relaxation process in a particular case.

THE great development in the technique of high polymers has brought a renewed interest in the theory of elastic and dielectric relaxation phenomena. The phenomenological theory of these effects is still based upon the classical work of Boltzmann, Hopkinson, and Volterra. Its recent advance has been made possible by the improvements of the mathematical tools—the Fourier and Laplace transforms, Stieltjes' integral equation, and Heaviside's operational calculus. The classic work on the elastic effect has been carefully reviewed by Leaderman.[1] Modern methods of specification and correlation of the properties of viscoelastic materials were discussed by Alfrey and Doty.[2] The present paper continues in the sense of the phenomenological theory. It treats the relations existing between the creep and the relaxation function, shows how to deduce from them the underlying more fundamental quantities—the distribution functions of retardation time of strain and of relaxation times of stress—and establishes a general transformation formula between these 2 distribution functions.

Permanent deformations shall not be taken into account, because here we are interested only in reversible effects. This does not constitute a serious limitation for the theory; it is possible to deduce from the measured deformation- and load-time curves the corresponding curves, which would be observed, if permanent deformations were absent. Furthermore our considerations are confined to the linear problem. They are entirely based on the assumption that the principle of superposition is valid for creep and for relaxation. On this basis the fundamental relation between the creep and the relaxation function is established.

Even the linear elastic system, under constant strain, does not in general follow the simple exponential law of stress decay, that on theoretical grounds should be expected for a homogeneous material. Soon the hypothesis was formulated that the macroscopic body, that constitutes the object for the measurement, possesses a complex structure, being built up by a multiplicity of different elementary systems, each of them obeying the exponential law and characterized by a proper time constant. The observed curve of decay then represents an integral effect, the linear superposition of the single exponential functions. The distribution function of relaxation times is then more characteristic for a substance than the observed relaxation curve. Wiechert[3] proposed for the distribution function a Gaussian curve; Wagner[4] used the same expression in his theory of the electric after-effect. But *a priori* there seems to be no reason to believe in the existence of a single general law for the distribution function; this function may well have different forms for different materials. Thus the problem arises to deduce the distribution function from the experimental creep and relaxation data. Fuoss and Kirkwood[5] have given an extremely interesting method that allows calcula-

---

* Publication assisted by the Committee on Inter-American Scientific Publication.

[1] H. Leaderman, *Elastic and Creep Properties of Filamentous Materials and Other High Polymers* (The Textile Foundation, Washington, D. C., 1944).

[2] T. Alfrey and P. Doty, J. App. Phys. 16, 700 (1945).

[3] E. Wiechert, Ann. d. Physik 50, 335 (1893).

[4] K. W. Wagner, Ann. d. Physik 40, 817 (1913).

[5] R. Fuoss and J. G. Kirkwood, J. Am. Chem. Soc. 63, 385 (1941).

tion of the distribution function of relaxation times, when the admittance function or its real or imaginary part are known. In the field of elastic phenomena, admittance and impedance measurements are not made with the same ease as for dielectrics and they would not lead to the determination of the long time end of the distribution function. Thus one must rely on the interpretation of long time creep and relaxation data. Simha[6] has already shown how to deduce the distribution function of relaxation times, when the creep function is given. Here this problem is treated by a method that is similar to and was suggested by that of Fuoss-Kirkwood; it leads to expressions for the distribution function, which lend themselves readily for analytic or numerical computation.

Parallel with stress relaxation under constant strain goes strain retardation under constant stress. And instead of the first, one can choose the second effect as the adequate basis for a discussion of the structure of the viscoelastic body. Then a reasoning analogous to that outlined above leads to the introduction of the distribution function of retardation times of strain; this derives from the creep function in the same way as does the other distribution from the relaxation function. Thus the problem of its determination is the same and is solved in the same way as the former one.

The two distribution functions are not independent from one another. The present theory establishes the transformation formulae for their mutual conversion. They are expressions which are of a very general scope and well suited for numerical computation.

The results of the general theory are applied to the study of the relaxation process in the particular case in which the creep function is represented by a given analytic expression. To start with, the simple power law is chosen. The corresponding relaxation function is found to be given by a known transcendent, the Mittag-Leffler function, which already has been discussed in detail in connection with various relaxation phenomena. It behaves reasonably at medium values, but does not describe correctly the behavior of the stress for very great values of time, giving too high a value for the amount of stress relaxation. To amend it, an amplitude factor is introduced in the law for the relaxation function, without a modification of the form of the law. With this new expression one calculates backward the creep function, to which it belongs, and gets in this way adequate representations for both the creep and the relaxation function. Then the distribution functions are calculated, and finally the behavior for small times is studied.

## I. RELATIONS BETWEEN THE CREEP AND THE RELAXATION FUNCTION

### Creep

The deformation, which in a viscoelastic body is produced by the application of a constant load $P_0$, increases with time. It contains a purely elastic, constant component $P_0F/E$, $E$ being the instantaneous value of the modulus of elasticity and $F$ a form factor, and a time-dependent, slowly increasing component $P_0\psi(t)F/E$. $\psi(t)$ is called the creep function. The total deformation under constant load is, therefore, given by

$$y(t)=P_0[1+\psi(t)]F/E. \tag{1}$$

The deformation produced by the application of a variable load $P(t)$ can be calculated with the aid of the principle of superposition

$$y(t)=\frac{F}{E}\left[P(t)+\int_{-\infty}^{t}P(\tau)\varphi(t-\tau)d\tau\right], \tag{2}$$

where

$$\varphi(t)=d\psi(t)/dt \text{ (rate of creep).} \tag{3}$$

### Relaxation

The load $P(t)$ necessary to produce in a viscoelastic body a constant deformation $y_0$ decreases with time, because the creep of the system gives rise to a phenomenon of stress-relaxation. There is

$$P(t)=y_0[1-\bar{\psi}(t)]E/F. \tag{4}$$

$\bar{\psi}(t)$ is the relaxation function.

The load necessary to produce a variable deformation $y(t)$ can again be calculated with the aid of the principle of superposition

$$P(t)=\frac{E}{F}\left[y(t)-\int_{-\infty}^{t}y(\tau)\bar{\varphi}(t-\tau)d\tau\right], \tag{5}$$

---

[6] R. Simha, J. App. Phys. 13, 201 (1942).

where

$$\bar{\varphi}(t)=d\bar{\psi}(t)/dt \text{ (rate of relaxation).}^7 \tag{6}$$

### Relations between Creep and Relaxation

The Eqs. (2) and (5) are 2 equivalent expressions of the fundamental experimental fact, on which this theory is based—the validity of the principle of superposition. Their coexistence implies a relation between the creep and the relaxation function. To obtain it easily, substitute $P(t)=\exp(pt)$ in (2) and (5). Since

$$\int_{-\infty}^{t} e^{p\tau}\varphi(t-\tau)d\tau=e^{pt}\int_{0}^{\infty} e^{-p\tau}\varphi(\tau)d\tau, \tag{7}$$

it follows that $y(t)=\text{const.}\exp(pt)$. There results finally

$$\bar{L}(p)+\bar{L}(p)L(p)-L(p)=0, \tag{8}$$

where

$$L(p)=\int_{0}^{\infty} e^{-pt}\varphi(t)dt, \tag{9a}$$

$$\bar{L}(p)=\int_{0}^{\infty} e^{-pt}\bar{\varphi}(t)dt. \tag{9b}$$

The deduction shows that $p$ may be real positive or complex, but it must not be on the negative real axis, because then the integrals would diverge. If $p$ is purely imaginary, equal to say $i\omega$, $L(i\omega)$ stands for the Fourier integral; if $p$ is real positive, $L(p)$ stands for the Laplace integral. Accordingly, 2 equivalent treatments are possible, i.e., by Fourier transforms or by Laplace transforms. Here we consider only the latter one.

Equation (8) is the fundamental relation, on which depend most of the following considerations. It may be solved with respect to $\bar{L}$ or to $L$:

$$\bar{L}(p)=\frac{L(p)}{1+L(p)}, \tag{10a}$$

$$L(p)=\frac{\bar{L}(p)}{1-\bar{L}(p)}. \tag{10b}$$

With the aid of these equations, it is possible to calculate the rate of creep, when the rate of relaxation is given, or *vice versa*.[8] In both cases, the wanted function is given by an integral equation—the Laplace transform. This must be inverted, when one wishes to obtain the explicit expressions for $\bar{\varphi}$ or $\varphi$. The inversion of the Laplace transform has recently been studied extensively.[9] A very general inversion formula now available is the Bromwich-Wagner or complex inversion formula; this was used by Simha[6] in his calculation of the distribution function of relaxation times. But other methods too may be employed, for instance, those of the operational calculus. And frequently one may find the solution with the aid of a table of Laplace transforms. Therefore, there is here no necessity to write down explicitly the formal expressions for the functions $\varphi$ and $\bar{\varphi}$, so more so as in the present case a more simple way of analysis is possible, based on the implicit relations (10). This is possible, because the functions we are dealing with are very particular ones; they are continuous, continuously decreasing and integrable functions, which in their turn can be represented as Laplace transforms.

## II. THE DISTRIBUTION FUNCTIONS FOR CREEP AND FOR RELAXATION

There are 2 fundamental representations of a viscoelastic body by way of models; the Voigt model and the Maxwell model. Alfrey and Doty[2] have given very recently in this Journal a concise discussion of the properties of these models. We, therefore, can confine ourselves to a summary description.

### The Voigt Model

The Voigt element consists of a spring and a dashpot connected in parallel. Creep under constant stress then follows an exponential law $\beta(1-\exp t/\tau)$, the time constant of which is the retardation time of strain. It indicates the time necessary for the time-dependent component of

---

[7] Different cases of creep, creep recovery, and relaxation as well as the deduction of the Eqs. (2) and (5) from (1) and (4) are discussed in detail by Leaderman (reference 1). The symbols used by this author are here conserved, but the functions $\varphi$ and $\bar{\varphi}$ are not normalized. This avoids the introduction of the 2 constants $\beta$ and $\bar{\beta}$, which in the present case are not relevant.

[8] Relations of this type are by no means new. For instance, they are already given by J. R. Carson, *Electric Circuit Theory and Operational Calculus* (McGraw-Hill Book Company, Inc., New York, 1926), pp. 51–52.

[9] See for instance, R. V. Churchill, *Modern Operational Methods in Engineering* (McGraw-Hill Book Company, Inc., New York, 1944).

deformation to reach a fraction $(1-e)$ of its final value. The Voigt model of a complex body consists of a multiplicity of Voigt elements with different time constants placed in series. The stress being the same for all, each element takes its share of the total strain. In the limit, the discontinuous system goes over in a continuous one, characterized by the distribution function $F(\tau)$ of retardation times. $F(\tau)d\tau$ is the fraction of the strain, which under constant stress has a retardation time between $\tau$ and $\tau+d\tau$. With the Voigt model, the fundamental experiment is the measurement of strain retardation, or creep, under constant stress. This gives the creep function, which in terms of the distribution function $F(\tau)$ is given by

$$\psi(t)=\beta\left[1-\int_0^\infty F(\tau)e^{-t/\tau}d\tau\right]; \qquad (11a)$$

$F(\tau)$ satisfies the normalizing condition

$$\int_0^\infty F(\tau)d\tau=1. \qquad (12)$$

### The Maxwell Model

The Maxwell element consists of a spring and a dashpot connected in series. Stress relaxation under constant strain follows an exponential law $\bar{\beta}(1-\exp t/\tau)$, the time constant of which is the relaxation time of stress. The Maxwell model of the complex body consists of a multiplicity of such elements placed in parallel. The strain being the same for all, each element takes its share of the stress, according to the value of its time constant. In the limit, the discontinuous system goes over in a continuous one, characterized by the distribution function $\bar{F}(\tau)$ of relaxation times. $\bar{F}(\tau)d\tau$ is the fraction of the stress, which under constant strain relaxes with a time constant between $\tau$ and $\tau+d\tau$. With the Maxwell model, the fundamental experiment is the measurement of stress relaxation under constant strain. This gives the relaxation function, which in terms of the distribution function $\bar{F}(\tau)$ is given by

$$\bar{\psi}(t)=\bar{\beta}\left[1-\int_0^\infty \bar{F}(\tau)e^{-t/\tau}d\tau\right]. \qquad (11b)$$

$\bar{F}$ satisfies a normalizing condition of the type (12).

### The Distribution Functions for Creep and for Relaxation

For the following it is more convenient to work with the functions $\varphi$ and $\bar{\varphi}$. It is easily seen, that these can be represented as

$$\varphi(t)=\int_0^\infty \rho(s)e^{-ts}ds, \qquad (13a)$$

$$\bar{\varphi}(t)=\int_0^\infty \bar{\rho}(s)e^{-ts}ds. \qquad (13b)$$

The functions $\rho$ and $\bar{\rho}$ introduced in this way are related to $F$ and $\bar{F}$ by the equations

$$\beta F(\tau)d\tau=-\rho(s)ds/s, \qquad (14a)$$

$$\bar{\beta}\bar{F}(\tau)d\tau=-\bar{\rho}(s)ds/s, \qquad (14b)$$

$$s=1/\tau. \qquad (14c)$$

We call $\rho$ the distribution function for creep and $\bar{\rho}$ the distribution function for relaxation. By means of Eqs. (14) it is always easy to go over to the "true" distribution functions $F$ and $\bar{F}$. The normalizing constants $\beta$ are determined with the aid of Eq. (12). Equations (14) show that the distribution functions can indeed be represented as Laplace transforms of "primitive" functions $\rho$. The possibility of this representation does not depend on the validity of the 2 models we have discussed, but ensues from the mathematical properties of the functions $\varphi$ and $\bar{\varphi}$. It is a common feature of all linear theories of relaxation phenomena.

## III. DETERMINATION OF THE DISTRIBUTION FUNCTIONS

Experimental evidence gives the functions $\varphi$ and $\bar{\varphi}$. Physically more interesting are the distribution functions $\rho$ and $\bar{\rho}$, because their knowledge may yield information about the structure of the system and the nature of the relaxation mechanism. One way of analysis consists in writing down an analytic expression for the distribution functions, which on theoretical grounds appears plausible and allows the calculation of the creep or the relaxation function, and then the comparison of the calculated and the measured curves. A second method consists in deducing the distribution functions from the experimentally given set of data. This can be done by an inversion of Eqs. (13).

## Distribution Function for Relaxation from Relaxation Function. Distribution Function for Creep from Creep Function

To obtain the inversion formula in the form needed here multiply both sides of Eq. (13b) by $e^{-pt}$ and integrate from 0 to $\infty$. This gives

$$\int_0^\infty e^{-pt}\bar{\varphi}(t)dt=\int_0^\infty e^{-pt}dt\int_0^\infty \bar{\rho}(s)e^{-st}ds. \quad (15)$$

The left side, according to Eq. (9b), is equal to $\bar{L}(p)$. On the right side the order of integration can be inverted and the integral over the exponential functions can be evaluated. This gives

$$\bar{L}(p)=\int_0^\infty \frac{\bar{\rho}(s)}{s+p}ds. \quad (16)$$

It is again not necessary for $p$ to be real; it may be complex, but it must not be on the negative real axis. Then (16) is Stieltjes' integral equation, which is inverted by[10]

$$\bar{\rho}(p)=\frac{1}{2\pi i}\lim_{\delta+0}[\bar{L}(-p-i\delta)-\bar{L}(-p+i\delta)]. \quad (17)$$

An analogous expression is obtained for $\rho(p)$.

To demonstrate the validity of (17), substitute in Eq. (16), $(-p-i\delta)$ for $p$. There follows

$$\lim_{\delta+0}\bar{L}(-p-i\delta)=\lim_{\delta+0}\int_0^\infty\frac{\bar{\rho}(s)(s-p)ds}{(s-p)^2+\delta^2}+\lim_{\delta+0}i\delta\int_0^\infty\frac{\bar{\rho}(s)ds}{(s-p)^2+\delta^2}. \quad (18)$$

In the first integral, one can put $\delta=0$, provided Cauchy's principal value is taken. To calculate the second integral, we observe that

$$\lim_{\delta+0}\delta\int_0^\infty\frac{\bar{\rho}(s)-\bar{\rho}(p)}{(s-p)^2+\delta^2}ds=0. \quad (19)$$

Indeed, the value of the integral *without* the factor $\delta$ in front of it is certainly finite even in the limit $\delta=0$; thus the expression *including* this factor disappears for $\delta=0$. But

$$\lim_{\delta+0}\delta\int_0^\infty\frac{\bar{\rho}(p)ds}{(s-p)^2+\delta^2}=\bar{\rho}(p)\lim_{\delta+0}\tan^{-1}(s/\delta)\Big|_{-p}^{\infty}=\pi\bar{\rho}(p). \quad (20)$$

The value of the second integral in (18) is, therefore, $i\pi\bar{\rho}(p)$. And

$$\lim_{\delta+0}\bar{L}(-p-i\delta)=\int_0^\infty\frac{\bar{\rho}(s)ds}{(s-p)}+i\pi\bar{\rho}(p); \quad (21a)$$

analogously,

$$\lim_{\delta+0}\bar{L}(-p+i\delta)=\int_0^\infty\frac{\rho(s)ds}{(s-p)}-i\pi\rho(p). \quad (21b)$$

The difference of both expressions gives $2\pi i\bar{\rho}(p)$.[11]

Equations (21) show that $\lim_{\delta+0}\bar{L}(-p-i\delta)$ and $\lim_{\delta+0}\bar{L}(-p+i\delta)$ are conjugate functions:

$$\lim_{\delta+0}[\bar{L}(-p-i\delta)-\bar{L}(-p+i\delta)]=2i\,Im\lim_{\delta+0}\bar{L}(-p-i\delta), \quad (22a)$$

$$\lim_{\delta+0}[\bar{L}(-p-i\delta)+\bar{L}(-p+i\delta)]=2\,Re\lim_{\delta+0}\bar{L}(-p-i\delta). \quad (22b)$$

$\lim_{\delta+0}\bar{L}(-p-i\delta)$ is equivalent to $\lim_{\epsilon+0}\bar{L}(\bar{p}e^{-i(\pi-\epsilon)})$; in a shorthand way this is written as $\bar{L}(pe^{-i\pi})$, provided the right meaning is attached to this expression: given $\bar{L}(p)$, $p$ positive; take the value that this function assumes when the argument is rotated through the under half-plan by an angle $-\pi$.[12] Accordingly, we replace the expressions $Im\lim_{\delta+0}\bar{L}(-p-i\delta)$ and $Re\lim_{\delta+0}\bar{L}(-p-i\delta)$, respectively, by $Im\,\bar{L}(pe^{-i\pi})$ and $Re\,\bar{L}(pe^{-i\pi})$. There results finally

$$\bar{\rho}(p)=Im\,\bar{L}(pe^{-i\pi})/\pi, \quad (23a)$$

$$\rho(p)=Im\,L(pe^{-i\pi})/\pi. \quad (23b)$$

The second equation is obtained by a development identical with that given above. When $\varphi$ and $\bar{\varphi}$ are known, $L$ and $\bar{L}$ can be calculated from (9). Then the distribution functions are obtained in a simple way. The deduction shows that the variable $p$ in (23) is equivalent with the variable $s$ of the Eqs. (14). Thus $\rho(p)dp$ and $\bar{\rho}(p)dp$ are, respectively, identic with $\rho(s)ds$ and $\bar{\rho}(s)ds$.

---

[10] T. J. Stieltjes, "Recherches sur les fractions continues," Ann. de la faculté de Toulouse 8, 1–22 (1894). Cf. D. V. Widder, *The Laplace Transform* (Princeton, University Press, New Jersey, 1941), p. 338. Am. Math. Mo. 52, No. 8 (1945).

[11] The evaluation of lim $\bar{L}(-p-i\delta)$ is a particular case of the calculus of the boundary values of a function along a given path. For a general solution of this problem see A. Hurwitz and R. Courant, *Funktionentheorie* (Verlagsbuchhandlung, Julius Springer, Berlin, 1929), pp. 333–335.

[12] Employing this terminology we follow E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Clarendon Press, Oxford, 1937), pp. 317–319, where a discussion of Stieltjes' integral equation is given.

### Distribution Function for Relaxation from Creep Function. Distribution Function for Creep from Relaxation Function

Frequently another type of relation is of interest. From the experimental point of view it is easier to measure the creep function than the relaxation function; but for theoretical reasons one may want the distribution function of relaxation times. This demands a formula that allows deduction of the distribution function for relaxation from the creep function, a problem that has already been discussed by Simha.[6] The particular type of inversion formula (17) makes it easy to solve this problem. By means of this equation, $\bar{\rho}(p)$ is expressed in function of the Laplace transform $\bar{L}(p)$ of $\bar{\varphi}(t)$; this in its turn is given in function of the Laplace transform $L(p)$ of $\varphi(t)$ by (10a). Substitution of (10a) in (17) directly gives, therefore, the wanted relation

$$\bar{\rho}(p)=\frac{1}{2\pi i}\lim_{\delta+0}\left[\frac{L(-p-i\delta)}{1+L(-p-i\delta)}-\frac{L(-p+i\delta)}{1+L(-p+i\delta)}\right]=\frac{1}{2\pi i}\frac{\lim\limits_{\delta+0}L(-p-i\delta)-\lim\limits_{\delta+0}L(-p+i\delta)}{[1+\lim\limits_{\delta+0}L(-p-i\delta)][1+\lim\limits_{\delta+0}L(-p+i\delta)]}. \quad (24)$$

An analogous expression is obtained for $\rho(p)$. Using the relations (22), the formula is transformed in a way similar to that of Eq. (17). There results after simple algebraic transformations

$$\bar{\rho}(p)=-\frac{1}{\pi}\frac{Im\,L(pe^{-i\pi})}{[1+Re\,L(pe^{-i\pi})]^2+[Im\,L(pe^{-i\pi})]^2}, \quad (25a)$$

$$\rho(p)=-\frac{1}{\pi}\frac{Im\,\bar{L}(pe^{-i\pi})}{[1-Re\,\bar{L}(pe^{-i\pi})]^2+[Im\,\bar{L}(pe^{-i\pi})]^2}. \quad (25b)$$

### An Auxiliary Formula

Equations (23) and (25) give the distribution functions in terms of the Laplace transforms of $\varphi$ and $\bar{\varphi}$, not in terms of the creep and relaxation functions themselves. These Laplace transforms are given by (9); when their values $L$ and $\bar{L}$ are known explicitly, the expressions $Im\,L(pe^{-i\pi})$ and $Re\,L(pe^{-i\pi})$ are easily calculated. For many functions, which are here of interest, the integrals can be evaluated in closed form. But in other cases this may not be so. Formulae, which allow to calculate the distribution functions without the previous explicit knowledge of $L(p)$ and $\bar{L}(p)$ are, therefore, of importance. Their deduction shall be discussed in a mathematical paper to be published elsewhere. The result is

$$Im\,L(pe^{-i\pi})=\int_0^\infty e^{-pt}\,Im\,\varphi(te^{-i\pi})dt, \quad (26a)$$

$$Re\,L(pe^{-i\pi})=-\int_0^\infty e^{-pt}\,Re\,\varphi(te^{-i\pi})dt. \quad (26b)$$

The same relations hold for $\bar{L}$ and $\bar{\varphi}$. They apply, when $\varphi$ in its turn is a Laplace transform.

For the application, the experimentally given set of creep or relaxation data must be represented by appropriate analytic functions, and the integrals (26) be calculated. Their evaluation may not always be possible in closed form, but provided the integrand can be written down explicitly, numerical methods can be employed.

## IV. RELATIONS BETWEEN THE DISTRIBUTION FUNCTIONS

The existence of a relationship between the creep and the relaxation function obviously also implies a direct relation between the corresponding distribution functions. For instance, assuming that the distribution function of relaxation times is represented by Wiechert's function, it must be possible to calculate the corresponding distribution function of retardation times, without previously calculating the creep and the relaxation function. To obtain this relation in a general form, we start with Eq. (24). The functions $\lim\limits_{\delta+0}L(-p\mp i\delta)$ are expressed in terms of $\rho(p)$ by means of (21). There results after simple algebraic transformations

$$\bar{\rho}(p)dp=\frac{\rho(p)dp}{\left[1+\int_0^\infty \frac{\rho(s)ds}{(s-p)}\right]^2+\pi^2\rho^2(p)}, \tag{27a}$$

$$\rho(p)dp=\frac{\bar{\rho}(p)dp}{\left[1-\int_0^\infty \frac{\bar{\rho}(s)ds}{(s-p)}\right]^2+\pi^2\bar{\rho}^2(p)}. \tag{27b}$$

These formulas convert one distribution function into another. They are of a very general scope, well adapted for analytic and numerical computation. Since the integration can be performed by graphical methods it would not even be necessary to represent $\rho$ or $\bar{\rho}$ by an analytic expression. The integrals are principal values, but the singularity at $s=p$ creates no difficulty; the way such integrals are handled has already been discussed in a former paper.[13]

## V. STUDY OF PARTICULAR CREEP AND RELAXATION FUNCTIONS

The simple power law

$$\psi(t)=kt^m, \quad 0<m<1 \tag{28}$$

is one of the functions which most frequently have been proposed as a mathematical expression for experimental creep data. It has the serious fault that it does not reach asymptotically a constant value for very great values of $t$, but yields an ever increasing deformation. But the fact alone that it has survived the many objections raised against it speaks much in favor of it being at least a reasonable approximation to the true fact, valid over a certain interval of time. Thus it is worth while to study the relaxation function that according to this theory is associated with it. The result shall lead us to a generalization of this law that is free of the fault mentioned above.

### The Simple Power Law

The rate of creep is given by

$$\varphi(t)=\beta(t/t_0)^{m-1}. \tag{29}$$

For the following it is convenient to introduce a numerical factor

$$\lambda=\beta t_0\Gamma(m), \tag{30}$$

where $\Gamma(m)$ is the gamma-function $\Gamma(m+1)=m!$[14] Substitution of (29) in (9a) gives

$$L(p)=\lambda(t_0p)^{-m}. \tag{31}$$

Then, according to Eq. (10a), $\bar{L}(p)$ is given by

$$\bar{L}(p)=\frac{\lambda(t_0p)^{-m}}{1+\lambda(t_0p)^{-m}}. \tag{32}$$

To obtain $\bar{\varphi}(t)$, the Laplace transform (9b) must be inverted, with $L(p)$ given by (32). There results

$$\bar{\varphi}(t)=-\frac{d}{dt}E_m[-\lambda(t/t_0)^m]. \tag{33}$$

$E_m(t)$ is a known transcendent, the function of Mittag-Leffler, defined by

$$E_m(t)=\sum_{n=0}^{\infty}\frac{t^n}{\Gamma(nm+1)}. \tag{34}$$

The relaxation function is given by

$$\bar{\psi}(t)=1-E_m[-\lambda(t/t_0)^m]. \tag{35}$$

*The deduction* of (33) is easily obtained by the Heaviside operational calculus.[15] Expand $pL(p)$ in descending powers of $p$. For the sake of simplicity put $\sigma=\lambda t_0^{-m}$. Then

$$p\bar{L}(p)=\sigma p^{1-m}[1-\sigma p^{-m}+\sigma^2p^{-2m}-\sigma^3p^{-3m}+-\cdots]. \tag{36}$$

Replace, according to the Heaviside rule, $1/p^n$ by $t^n/\Gamma(1+n)$. This gives

$$\bar{\varphi}(t)=\frac{\sigma t^{m-1}}{\Gamma(m)}-\frac{\sigma^2t^{2m-1}}{\Gamma(2m)}+\frac{\sigma^3t^{3m-1}}{\Gamma(3m)}-\frac{\sigma^4t^{4m-1}}{\Gamma(4m)}+-\cdots. \tag{37}$$

Observing that $m\Gamma(m)=\Gamma(m+1)$, Eq. (37) is easily identified with the series expansion of the derivative of the Mittag-Leffler function of argument $\{-\lambda(t/t_0)^m\}$. The inversion of the Laplace transform in the present case can also be obtained by the more orthodox method of the complex inversion theorem. This was shown by Moraes and Schoenberg.[16]

An *asymptotic expansion* valid for great values of $t/t_0$ is also obtained in the Heaviside way. Expand $pL(p)$ in ascending powers of $p$.

$$p\bar{L}(p)=p[1-p^m/\sigma+p^{2m}/\sigma^2-p^{3m}/\sigma^3+p^{4m}/\sigma^4-+\cdots]. \tag{38}$$

---

[13] H. Silva and B. Gross, Phys. Rev. 60, 684 (1941). The method indicated in this paper has recently been used by S. T. Ma, Phys. Rev. 68, 166 (1945).

[14] Tables and integral representations of the gamma-function are given by Jahnke-Emde, *Tables of Functions* (B. Teubner, Leipzig, 1933), pp. 86, 96.

[15] Reference 8, p. 28. $\bar{L}(p)$ of the present paper corresponds to Carson's $1/pH(p)$; thus Carson's expansion of $1/H(p)$ corresponds in our case to an expansion of $p\bar{L}(p)$.

[16] A. de Moraes and M. Schoenberg, Ann. Acad. Brasil. Ci. 12, 137 (1940), Section 4.

Replace $p^n$ by $t^{-n}/\Gamma(1-n)$, discard the term with an integer power in $p$ (because $p^1$ gives $1/t\Gamma(0)$; but $\Gamma(0)=\infty$) and get

$$\bar{\varphi}(t)=-\frac{\sigma^{-1}t^{-m-1}}{\Gamma(-m)}+\frac{\sigma^{-2}t^{-2-m1}}{\Gamma(-2m)}-\frac{\sigma^{-3}t^{-3m-1}}{\Gamma(-3m)}+\frac{\sigma^{-4}t^{-4m-1}}{\Gamma(-4m)}-+\cdots. \quad (39)$$

*The Mittag-Leffler function* is a generalization of the exponential function, with which it is identical, when the parameter $m$ is 1. It is already known to be of importance for relaxation phenomena of various kinds. Several of its mathematical aspects as well as its connection with the theory of nerve conduction have been discussed by Davis.[17] In connection with the theory of the discharge and return voltage curves for absorptive capacitors it has been studied by Castro, who has also availed himself of a very useful approximate representation.[18] For certain integer values of $m$ it can be expressed by elementary transcendents. For small values of $m$ a simple approximation formula exists. There is furthermore $E_m(0)=1$ and $E_m(-\infty)=0$.

Very valuable information about the behavior of the expression (33) is contained in a paper by Cole and Cole.[19] A comparison of their formulae (8) and (10) with the Eqs. (37) and (39) of the present paper shows them to be identical, up to a constant, the parameter $m$ corresponding to Cole's $1-\alpha$. (The Coles write $-nm\Gamma(1-nm)$ for $\Gamma(-nm)$.) This shows, that the Cole's expression for the absorption current is identic with our expression (33) for the rate of relaxation; it allows to use Fig. 1 and Table I of the Cole's paper for computation of the rate of relaxation in exactly the same way as those authors use them for the computation of the dielectric absorption current, and relieves us here of the task to give a plot or tabulate the function $\bar{\varphi}(t)$.

For very small values of $t$, $\bar{\varphi}(t)$ is given by the leading term of (37). Thus, when $t$ approaches 0, $\bar{\varphi}(t)$ tends to $\infty$ in the same way as does $\varphi(t)$

$$t\rightarrow 0:\quad \bar{\varphi}(t)=\bar{\beta}(t/t_0)^{m-1}. \quad (40)$$

For very great values of $t$, $\bar{\varphi}(t)$ is given by the leading term of (39). Observing that $\Gamma(m)\Gamma(-m)=-\pi/m\sin\pi m$,[14] there results

$$t\rightarrow\infty:\quad \bar{\varphi}(t)=\frac{m\sin\pi m}{\pi\beta t_0{}^2}(t/t_0)^{-(m+1)}. \quad (41)$$

Thus when $t$ approaches $\infty$, $\bar{\varphi}(t)$ tends to 0 much more rapidly than does $\varphi(t)$. It follows that the integral taken from 0 to $\infty$ remains finite.

It is known that experimental creep data frequently call for a very small value of $m$. In this case a satisfactory elementary approximation for $E_m(t)$ exists.[18] It allows to write $\bar{\varphi}(t)$ in the form

$$\bar{\varphi}(t)=\beta(t/t_0)^{m-1}\frac{1}{[1+\beta t_0(t/t_0)^m/m]^2}, \quad (42)$$

an expression that shows well the general behavior of the function and its departure from the simple power law.

## The Mittag-Leffler Function

It cannot be expected that these formulae describe correctly the experimental relaxation data for great values of $t$, based as they are on a creep law that admittedly fails for such values. Indeed, an inspection of Eq. (35) shows that $\bar{\psi}(t)$ approaches 1, when $t$ approaches $\infty$. This means, that in the case of stress relaxation under constant strain (Eq. (4)) the stress decays to 0, an absurd result that obviously is a consequence of the infinite increase of the creep function (28). In reality the amount of stress relaxation is only a rather small fraction of the total stress. To get a relaxation function that fulfills this condition it is not necessary to modify the general form of the relaxation law (35). It is sufficient to postulate that the amplitude of the function is not 1, but smaller than 1. Thus one must introduce an amplitude factor $a<1$, writing

$$\bar{\psi}(t)=a[1-E_m\{-(t/t_0)^m\}]. \quad (43)$$

A comparison with Eq. (4) shows the significance of $a$; $1-a$ represents the relation between final load $P(\infty)$ and initial load (or stress) $P(0)$.

$$1-a=P(\infty)/P(0). \quad (44)$$

We, therefore, are led to believe that Eq. (43) is an adequate analytic expression for the relaxation function. Now one has to calculate the creep function to which it belongs.

The rate of relaxation is now given by

$$\bar{\varphi}(t)=-a\frac{d}{dt}E_m[-(t/t_0)^m]. \quad (45)$$

The corresponding Laplace transform $\bar{L}(p)$, according to (32) and (33), is

$$\bar{L}(p)=a\frac{(t_0p)^{-m}}{1+(t_0p)^{-m}}. \quad (46)$$

---

[17] H. T. Davis, *The Theory of Linear Operators* (The Principia Press, Bloomington, 1936).

[18] F. M. de Oliveira Castro, Zeits. f. Physik **114**, 116 (1939).

[19] K. S. Cole and R. H. Cole, J. Chem. Phys. **10**, 98 (1942).

Substitution of (46) in (10b) gives

$$L(p)=\frac{a}{1-a}\frac{(1-a)(t_0p)^{-m}}{1+(1-a)(t_0p)^{-m}}. \qquad (47)$$

Therefore,

$$\varphi(t)=-\frac{a}{1-a}\frac{d}{dt}E_m[-(1-a)(t/t_0)^m]. \qquad (48)$$

Both the creep and the relaxation function are now represented by the same general law; they differ only by the values of the parameters. This is very satisfactory; it brings out the experimental fact that those functions are of the same type. By an appropriate choice of $a$, $t_0$, and $m$, a great variety of experimental curves can be represented, varying from the simple power law to the exponential function. The asymptotic expressions (40) and (41) and the approximation formula (42) now apply to creep and to relaxation; they must only be completed by the corresponding amplitude factor. The numerical analysis of creep and relaxation data can be performed by the methods given by Cole and Cole[19] and already mentioned above. But it cannot be the purpose of this paper to enter into details of numerical computation.

The problem could have been approached from the side of the creep function too. The expression (33) is in itself a generalization of the simple power law, coinciding with it for small values, but converging more rapidly at great values, of $t$. This would suggest the possibility of introducing it as a suitable substitute for the power law. In the corresponding electric case, Cole and Cole[19] proceeded in an analogous way, when they represented the dielectric absorption current by such an expression.

## A Functional Relation between Creep and Relaxation Function

Since $\varphi(t)$ and $\bar{\varphi}(t)$ are given by the same general law, a direct relation between them is readily established. It follows from Eqs. (48) and (45) that

$$\bar{\varphi}(t)=(1-a)^{(m-1)/m}\varphi\left[\frac{t}{(1-a)^{1/m}}\right]. \qquad (49)$$

This possibility of direct computation of relaxation data from creep measurements should be of particular interest for technical applications.

## The Distribution Function

The distribution function of relaxation times, that belongs to the relaxation function (43) is easily calculated. Since the Laplace transform $\bar{L}(p)$ of $\bar{\varphi}(t)$ is known explicitly, Eq. (23) applies directly. Replace in (46) $p^m$ by $(pe^{-i\pi})^m$. Then

$$\bar{L}(pe^{-i\pi})=\frac{a}{1+(t_0p)^m(\cos\pi m-i\sin\pi m)}. \qquad (50)$$

To get $\bar{\rho}(p)$ one has to take the imaginary part of this expression. Thus

$$\bar{\rho}(p)dp=-\frac{1}{\pi}\frac{a\sin\pi m\,dp}{(t_0p)^m+(t_0p)^{-m}+2\cos\pi m}. \qquad (51)$$

The distribution function has its maximum at $p=1/t_0$. Introducing a new variable

$$u=\log_e 1/t_0p \qquad (52)$$

normalizing, and observing that $p=1/\tau$, one gets

$$\bar{F}(\tau)d\tau=\frac{1}{2\pi}\frac{\sin\pi m}{\cosh um+\cos\pi m}du. \qquad (53)$$

The distribution function of retardation times is given by the same law, but while for relaxation $u=\log_e\tau/t_0$, there is for creep

$$u=\log_e\{\tau/t_0(1-a)^{1/m}\}.$$

$\bar{F}(\tau)d\tau$ as given by Eq. (53) is easily identified with the distribution function that Cole and Cole[20] have obtained for dielectric relaxation by the Fuoss-Kirkwood method.[6] These authors have discussed in detail the behavior of the function and compared it with the Wiechert-Wagner distribution.

## The Behavior for Small Times

The rates of creep and of relaxation as given by the Eqs. (45) and (48) become infinite at $t=0$, while theoretical expectations let us expect them to be finite. The study of the behavior for

---

[20] K. S. Cole and R. H. Cole, J. Chem. Phys. **9**, 341 (1941). Coles' parameter $1-\alpha$ corresponds to the parameter $m$ of the present paper.

small times should, therefore, be carried out with a class of functions that remain finite for $t=0$. An adequate representation for this interval would be

$$\bar{\varphi}(t)=\beta(c+t/t_0)^{m-1}, \quad c\ll 1. \tag{54}$$

The Laplace transform of this function cannot be written down explicitly. Thus one has to rely on the Eqs. (26). Here we shall confine ourselves to the discussion of the distribution function for relaxation. To obtain it, replace in Eq. (54) $t$ by $te^{-i\pi}$ and take the imaginary part of the expression $\bar{\varphi}(te^{-i\pi})$. This gives

$$Im\ \bar{\varphi}(te^{-i\pi})=\begin{cases}0 & \text{for } t<ct_0,\\ \beta\sin\pi(1-m)\,(t/t_0-c)^{m-1} & \text{for } t>ct_0.\end{cases} \tag{55}$$

The expression (55) is substituted in (26a). The integrand being 0 for $t<ct_0$, the integral now runs from $ct_0$ to $\infty$. By a change of variable, the integral is reduced to a gamma-function; the final result is[21]

$$\bar{\rho}(p)dp=\frac{\beta t_0}{\pi}\Gamma(m)\,\sin\pi(1-m)\frac{e^{-c(t_0p)}}{(t_0p)^{n}}$$

This distribution must be compared with (51). Here we are concerned only with small values of $t/t_0$, i.e., small values of $\tau$. Since $p=1/\tau$, this means, that we must consider only values of $(t_0p)$ very much greater than $c$. For such values, the distribution (51) decreases like $1/(t_0p)^m$, while the new distribution (56) decreases like $e^{-c(t_0p)}/(t_0p)^m$, i.e., much more rapidly than the former one. Practically, this signifies an exponential cut-off of the distribution function for small relaxation times.

I am very indebted to the Director of the Institute, Professor E. L. da Fonseca Costa, who made this study possible.

---

[21] This result can be found in tables of Laplace transforms; see for instance, Carson (reference 8), p. 39, formula (e). Here the deduction is outlined, because it is a good example for the application of Eq. (26).

---

# Parallel Plate Optics for Rapid Scanning*

S. B. Myers**
*Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Received August 28, 1946)

Rapid scanning using a paraboloid antenna is unsatisfactory; rapid movement of the antenna is mechanically impractical, while motion of the feed near the focus requires long focal length for good optical results. Another attack on the problem is to place the feed between parallel plates which produce a line source near a bifocal reflector. The energy flows along the geodesics of the mean surface between the parallel plates; the problem of shaping the plates so that a circular feed motion produces an oscillating beam at the line source becomes a problem in differential geometry and the calculus of variations. If the mean surface is assumed to consist of two developable surfaces joined along part of their boundaries, the unique solution of the optical problem is a circular disk along the edge of which the feed moves, joined to a section of a 60° cone which is straightened out at the large end to obtain a straight line source. This is the RCA solution of the problem. If the surface is assumed to consist of a surface of revolution which provides a feed circle, and attached to it a circular disk a diameter of which is the aperture, under certain natural restrictions no exact solution of the optical problem is possible, but several approximate solutions exist. Dielectric introduced between the plates of such a surface of revolution scanner effects a great improvement in the optics.

## INTRODUCTION

RAPID scanning by means of movement of a paraboloid antenna is mechanically impractical. If the paraboloid is kept fixed, and the source of energy moved near the focus, the optical results are not satisfactory except with an extremely long focal length. Another approach to the problem is the following. If the energy feed be placed between two curved parallel plates, experiment has shown that the energy

---

* This paper is based on work done for the Office of Scientific Research and Development under Contract OEMsr-262 with the Massachusetts Institute of Technology.

** Now at the University of Michigan.

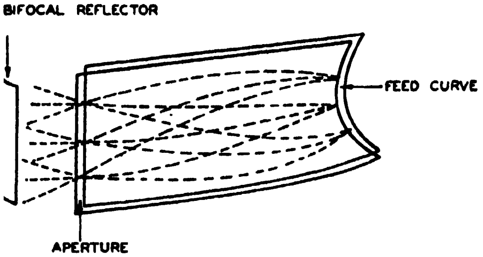


FIG. 1.

will flow along geodesics of the mean surface between the plates; if the parallel plates can be shaped so that a motion of the feed will produce an oscillating beam at a straight aperture, the latter will serve as a line source for a nearby bifocal reflector and the desired scanning is possible. For best illumination it is desirable that the normal ray from every point on the feed curve pass through a fixed point on the aperture; also, it is mechanically desirable that the feed curve be a circle. (See Fig. 1.)

A directed beam will be produced at the aperture by a fixed feed 0 if the phase errors at the aperture vary linearly with length along the aperture. The simplest way to accomplish this is to shape the parallel plates so that the mean surface is planar near the aperture, and the wave fronts are straight (near the aperture). This latter condition implies that the geodesics from 0 are parallel near the aperture. We are thus led to the following geometric problem. To find a surface $S$ containing a straight line $C_2$ and planar near $C_2$, and containing a curve $C_1$ (preferably a circle), with the property that the geodesics issuing from each point 0 of $C_1$ meet $C_2$ at a constant angle (which varies as 0 moves on $C_1$); further, the geodesics normal to $C_1$ must all meet on $C_2$.

This problem involves differential geometry and the calculus of variations. As yet it is not known whether a surface exists which meets all these requirements. Further, additional stringent restrictions as to size and shape of the antenna are usually imposed in any practical application. In this paper we discuss several approximate solutions of the problem, which seem suitable for application.

In Section 2 we take up a solution discovered by RCA; it yields perfect scanning optics (i.e., produces an oscillating beam at the aperture), but fails to meet the illumination requirement. The mean surface consists of a circular disk along the edge of which the feed moves, and a section of a 60° cone which is straightened out at the large end to obtain a straight line aperture (see Fig. 2). It is the only known surface which solves the optical problem exactly, and it is shown here that it is the unique method of joining two developable surfaces so as to obtain such optics. In Section 3, the "surface of revolution" solution is studied. Here the mean surface consists of a surface of revolution which provides a feed circle, and attached to it a circular disk a diameter of which is the aperture (see Fig. 3). Such a solution satisfies all requirements except the optical conditions. It is proved that these optical conditions cannot be satisfied exactly by any such surface if the radius of the surface is bounded by the radii of the feed circle and aperture disk. An investigation of the problem of minimizing the optical errors reveals that by shaping the surface in any one of several simple ways these errors can be made very small. In Section 4 it is shown that this same solution can be improved by use of dielectric between the parallel plates.

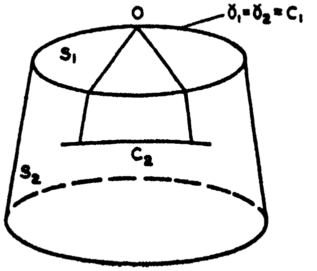


FIG. 2.

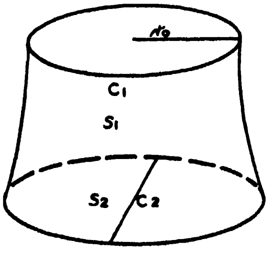


FIG. 3

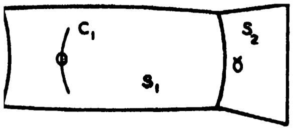


Fig. 4.

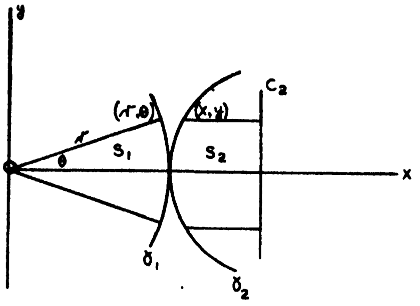


Fig. 5.

## 1. SURFACES AND GEODESICS[1]

Let a surface $S$ be given by equations $x=x(u, v)$, $y=y(u, v)$, $z=z(u, v)$. From the usual formula

$$ds^2=dx^2+dy^2+dz^2=A(u, v)du^2 +2B(u, v)dudv+C(u, v)dv^2$$

is obtained the expression for the length of a curve $u=u(t)$, $v=v(t)$.

$$s=\int(Au'^2+2Bu'v'+Cv'^2)^{\frac{1}{2}}dt =\int F(u, v, u', v')dt.$$

A geodesic may be defined as a shortest arc joining its endpoints. Using $s$ as parameter, the geodesics on $S$ are solutions of the equations

$$\frac{d}{ds}\left(\frac{\partial F}{\partial u'}\right)=\frac{\partial F}{\partial u},$$

$$F(u, v, u', v')=1.$$

Along a geodesic $g$, the geodesic curvature is zero (geodesic curvature of a curve $g$ at a point $P$ is by definition the curvature of the projection of $g$ on the tangent plane to $S$ at $P$).

Two fundamental properties of geodesics needed here are: (1) through every point on a surface and in every direction there is a unique geodesic, (2) if a constant length is measured off on geodesic rays through a point 0 the locus (geodesic circle) thus obtained is perpendicular to the rays through 0 (in other words, wave fronts from a point source are normal to the rays from the source). It can be shown that a geodesic crossing a sharp bend in a surface must make equal angles with the bend on both sides.

A developable surface is one which can be deformed locally onto a plane without stretching or tearing, thus preserving lengths. This is equivalent to the identical vanishing of the Gauss curvature. The fundamental form for a developable is the same as for the plane, $ds^2=du^2+dv^2$ or in polar coordinates, $dr^2+r^2d\theta^2$. Examples of developables are cylinders and cones. Two developables $S_1$ and $S_2$, $S_1$ containing a curve $C$ and $S_2$ containing a curve $C'$, can be joined along $C, C'$ as common curve without deforming $S_1$ if the curvature of $C$ at each point is greater than the geodesic curvature of $C'$ at the corresponding point. The geodesics on a developable become straight lines when the surface is developed onto a plane.

## 2. THE RCA SOLUTION

We take up the problem of finding a surface $S$, formed by joining two developables $S_1$ and $S_2$ along a common edge $\gamma$ (see Fig. 4), with the property that $S_1$ contains a curve $C_1$ and $S_2$ a straight line segment $C_2$ such that all geodesics from one point 0 of $C_1$ to the points of $C_2$ have equal lengths (this is equivalent to the condition that all these geodesics be perpendicular to $C_2$), and all geodesics from each point of $C_1$ near 0 to the points of $C_2$ have lengths which vary linearly with length along $C_2$ (this is equivalent to the condition that these geodesics meet $C_2$ at constant angle).

Let $S_1$ and $S_2$ be two developables joined along a common edge $\gamma$. Suppose each developable is developed into planar form, the curve $\gamma$ becoming two curves $\gamma_1$ and $\gamma_2$. Let 0 be a point in $S_1$ and let $\gamma_1$ and $\gamma_2$ be in contact at the foot of the normal from 0 to $\gamma_1$. Using coordinates as shown in Fig. 5, let the polar equation of $\gamma_1$ be $r=r(\theta)$. Now

$$ds^2=dr^2+r^2d\theta^2 \text{ along } \gamma_1,$$
$$ds^2=dx^2+dy^2 \text{ along } \gamma_2.$$

---

[1] See, for example, L. P. Eisenhart, *A Treatise on the Differential Geometry of Curves and Surfaces* (Ginn and Company, New York, 1909).

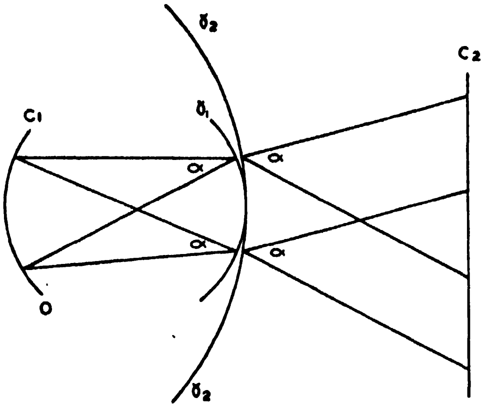


FIG. 6.

If the aperture $C_2$ is to be a wave front for rays from 0 when $\gamma_1$ and $\gamma_2$ are made to coincide, we must have $x=r(\theta)$ along $\gamma_2$. Hence $dx=dr$, $dy=rd\theta$ and the cartesian equations of $\gamma_2$ are

$$x=r(\theta),$$
$$y=\int_0^\theta r(\theta)d\theta. \tag{1}$$

If $C_2$ is to continue to cut the rays from the source 0 at equal angles as 0 moves along a feed curve $C_1$, then each pair of points on $C_1$ must subtend a constant angle $\alpha$ at all points of $\gamma$. Therefore $\gamma_1$ is a circle passing through each pair of points of $C_1$; hence $C_1$ and $\gamma_1$ are parts of the same circle $r=2a\cos\theta$. From (1) it follows that $\gamma_2$ has equations

$$x=2a\cos\theta$$
$$y=2a\sin\theta$$

and is a circle whose radius is double that of $\gamma_1$ (see Fig. 6).

If one-half of the circle $\gamma_2$ is made to coincide with $\gamma_1$, leaving $S_1$ fixed, the surface $S_2$ is deformed into the lateral surface of a cone inclined at a 60° angle to the circular disk $S_1$ (see Fig. 2). The aperture $C_2$ is now a curve on the side of a cone, but the figure can be deformed so that $C_2$ becomes straight in space. As the feed 0 rotates through a central angle $\beta$ on $C_1$, the beam issuing from $C_2$ tilts through an angle $\beta$.

This is the RCA solution. It furnishes the unique method of joining two developables so as to get exact scanning optics and is the only known method of obtaining such optics. However as 0 moves along $C_1$ the illumination pattern along $C_2$ changes, since the rays normal to $C_1$ do not pass through the same point of $C_2$.

### 3. SURFACE OF REVOLUTION SCANNER

A promising type of surface $S$ is composed of a surface of revolution $S_1$ with a base disk $S_2$ attached to it along a circular cross section (see Fig. 3). The feed curve $C_1$ is a circular cross section of $S_1$ and the aperture $C_2$ is the diameter of $S_2$. All the geodesics normal to $C_1$ pass through the center of the aperture, so that the illumination requirement is satisfied. Symmetry properties of $S$ show that if the geodesics from one point 0 of $C_1$ meet $C_2$ at a constant angle, so will the geodesics from any other point of $C_1$. We will show that such perfect optics are impossible if the radius of the surface is bounded by the radii of $C_1$ and $S_2$, and will investigate the question of shaping the surface so as to minimize the optical errors.

If $S_1$ is formed by rotating the curve $y=y(r)$ about the $y$ axis we obtain[2]

$$ds^2=dr^2+r^2d\theta^2+dy^2=(1+y'^2)dr^2+r^2d\theta^2,$$
$$y'=\frac{dy}{dr},$$
$$s=\int\left((1+y'^2)\left(\frac{dr}{dt}\right)^2+r^2\left(\frac{d\theta}{dt}\right)^2\right)^{\frac{1}{2}}dt.$$

The equations of geodesics on $S_1$ can be integrated once to give

$$r^2(d\theta/ds)=h,$$

where $h$ is a constant along each geodesic. But if

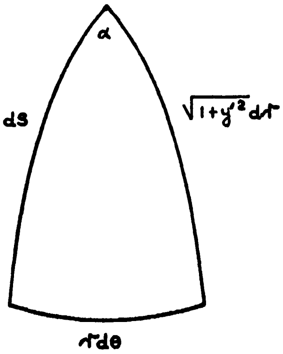


FIG. 7.

---

[2] Cf. Eisenhart, reference 1, pp. 205–206.

we denote by $\alpha$ the variable angle at which a geodesic meets a meridian, it is easy to see (Fig. 7) that

$$d\theta/ds=\sin\alpha/r.$$

Therefore along a geodesic $r\sin\alpha=h$. Since

$$\cos\alpha=(1+y'^2)^{\frac{1}{2}}\frac{dr}{ds},$$

we find that along a geodesic

$$r^2\left[1-(1+y'^2)\left(\frac{dr}{ds}\right)^2\right]=h^2.$$

Let 0 be the point on $C_1$ such that the plane through 0 and the axis of the surface is normal to the aperture $C_2$, let $g$ be a geodesic issuing from 0, and suppose that at 0 $r=r_0$ and $\theta=0$. Let $r_1$ be the radius of the base circle, $s_1$ the length of $g$ from 0 to the base circle, and $\alpha_0$, $\alpha_1$ the initial and final values of $\alpha$ on $g$ (see Fig. 8). Then

$$\left.\begin{aligned} s_1&=\left|\int_{r_0}^{r_1}\frac{(1+y'^2)^{\frac{1}{2}}r\,dr}{(r^2-h^2)^{\frac{1}{2}}}\right|,\\ \theta_1&=h\left|\int_{r_0}^{r_1}\frac{(1+y'^2)^{\frac{1}{2}}dr}{r(r^2-h^2)^{\frac{1}{2}}}\right|,\\ \alpha_1&=\arcsin\frac{h}{r_1}.\end{aligned}\right\}\qquad\text{(A)}$$

As $g$ enters the base disk, it becomes a straight line, whose perpendicular distance from the center of the aperture is $h$. Since the meridian geodesic from 0 meets the aperture at right angles, the problem of obtaining perfect optics is that of making all geodesics from 0 be normal to the aperture $C_2$; that is, we want $\theta_1=\alpha_1$ (see Fig. 9).

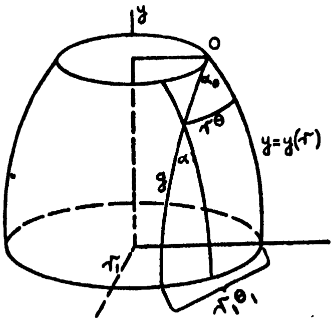


FIG. 8.

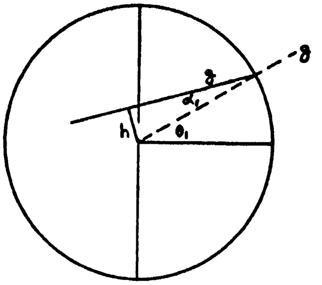


FIG. 9.

If $\theta_1-\alpha_1$ is expanded in a series about $h=0$, we obtain

$$\theta_1-\alpha_1=\sum_{n=1}^{\infty}A_{2n-1}\frac{h^{2n-1}}{2n-1}\frac{1\cdot3\cdot5\cdots(2n-1)}{2^{n-1}(n-1)!(2n-1)},$$

where

$$A_{2n-1}=(2n-1)\left|\int_{r_0}^{r_1}\frac{(1+y'^2)^{\frac{1}{2}}dr}{r^{2n}}\right|-\frac{1}{r_1^{2n-1}}.$$

We now prove the impossibility of $\theta_1=\alpha_1$ for all $h$ by showing that

$$A_{2n+1}>\frac{A_{2n-1}}{\bar{r}^2},$$

where $\bar{r}$ is the larger of $r_0$, $r_1$, so that it is impossible that $A_{2n-1}=0$ for all $n$.

Let $F(r)=(1+y'^2)^{\frac{1}{2}}$. If $r_1>r>r_0$,

$$A_{2n+1}=(2n+1)\int_{r_0}^{r_1}\frac{Fdr}{r^{2n+2}}-\frac{1}{r_1^{2n+1}}$$

$$>(2n-1)\int_{r_0}^{r_1}\frac{Fdr}{r_1^2r^{2n}}-\frac{1}{r_1^{2n+1}}=\frac{A_{2n-1}}{r_1^2}.$$

If $r_0>r>r_1$,

$$A_{2n+1}=(2n+1)\int_{r_1}^{r_0}\frac{Fdr}{r^{2n+2}}-\frac{1}{r_1^{2n+1}}$$

$$=(2n+1)\int_{r_1}^{r_0}\frac{(F-1)dr}{r^{2n+2}}-\frac{1}{r_0^{2n+1}}$$

$$>(2n-1)\int_{r_1}^{r_0}\frac{(F-1)dr}{r_0^2r^{2n}}-\frac{1}{r_0^{2n+1}}=\frac{A_{2n-1}}{r_0^2}.$$

This proof can also be carried through for the case of a contour curve of the more general type $r=r(t)$, $y=y(t)$ provided that the upper and lower radii $r_0$ and $r_1$ are extreme values of $r$.

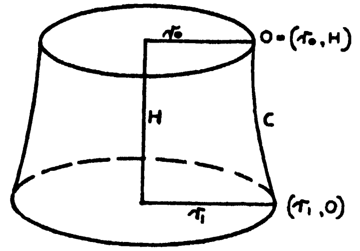


FIG. 10.

If $r_1>r_0$.

$$A_{2n+1}=(2n+1)\int_{t_0}^{t_1}\frac{\left[\left(\frac{dr}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2\right]^{\frac{1}{2}}dt}{r^{2n+2}}$$

$$\frac{1}{r_1^{2n+1}}>\frac{A_{2n-1}}{r_1^2}.$$

If $r_0>r_1$,

$$A_{2n+1}=(2n+1)\int_{t_0}^{t_1}\frac{\left[\left[\left(\frac{dr}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2\right]^{\frac{1}{2}}+\frac{dr}{dt}\right]dt}{r^{2n+2}}$$

$$\frac{1}{r_0^{2n+1}}>\frac{A_{2n-1}}{r_0^2}.$$

Thus we have proved the existence of phase errors at the aperture, provided that $r$ remains between $r_0$ and $r_1$ on the surface.

The size of these errors depends on the contour curve $C$, the upper and lower radii $r_0$ and $r_1$, and the height $H$. For fixed $r_1$, $H$ the phase errors approach zero as $r_0\rightarrow\infty$. But practical considerations restrict $r_0$, $r_1$, $H$. In view of applications the following problem is formulated.

For fixed $r_0$, $r_1$, $H$ what is the contour curve $C$ which maximizes the amount of the aperture for which the phase errors remain less than a pre-assigned amount?

The length of a geodesic from 0 to the aperture is easily computed to be

$$s_1+r_1\frac{\cos\theta_1}{\cos(\theta_1-\alpha_1)},$$

where $s_1$, $\theta_1$, $\alpha_1$ are given by Eqs. (A). The phase errors are given by the difference between this length and the length of the meridian ray from 0 to the aperture.

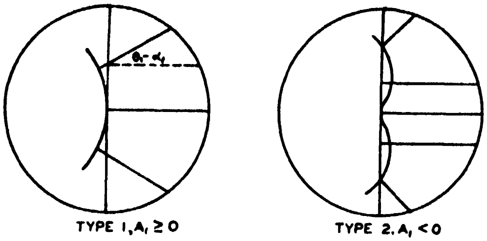


FIG. 11.

The maximizing problem stated above seems too difficult to approach directly, and we replace it by the problem of maximizing the amount of the aperture for which the angular error $|\theta_1-\alpha_1|$ remains less than a pre-assigned amount. This in turn we replace by the following first approximation (see Fig. 10).

What is the curve $C$: $y=y(r)$ in the $(r, y)$-plane which furnishes a minimum value to $A_3$ among all curves of the form $y=y(r)$ joining $(r_0, H)$ to $(r_1, 0)$ and making $A_1=0$? ($A_1$ and $A_3$ as before are the first terms occurring in the expansion of $\theta_1-\alpha_1$.)

The graph of a wave front is either of two types shown in Fig. 11, depending on whether $A_1\geqq 0$ or $A_1<0$.

The problem just formulated is equivalent geometrically to flattening the wave front of Type 1 at its center as much as possible. Analytically, we are led to an isoperimetric problem[3] in the calculus of variations—to find the curve $y=y(r)$ which belongs to the class of curves joining $(r_0, H)$ to $(r_1, 0)$ and making

$$\left|\int_{r_0}^{r_1}\frac{(1+y'^2)^{\frac{1}{2}}dr}{r^2}\right|-\frac{1}{r_1}=0,$$

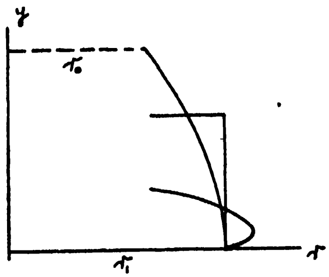


FIG. 12.

---

[3] See, for example, O. Bolza, *Vorlesungen über Variationsrechnung* (B. G. Teubner, Leipzig, 1933), Chapter 10.

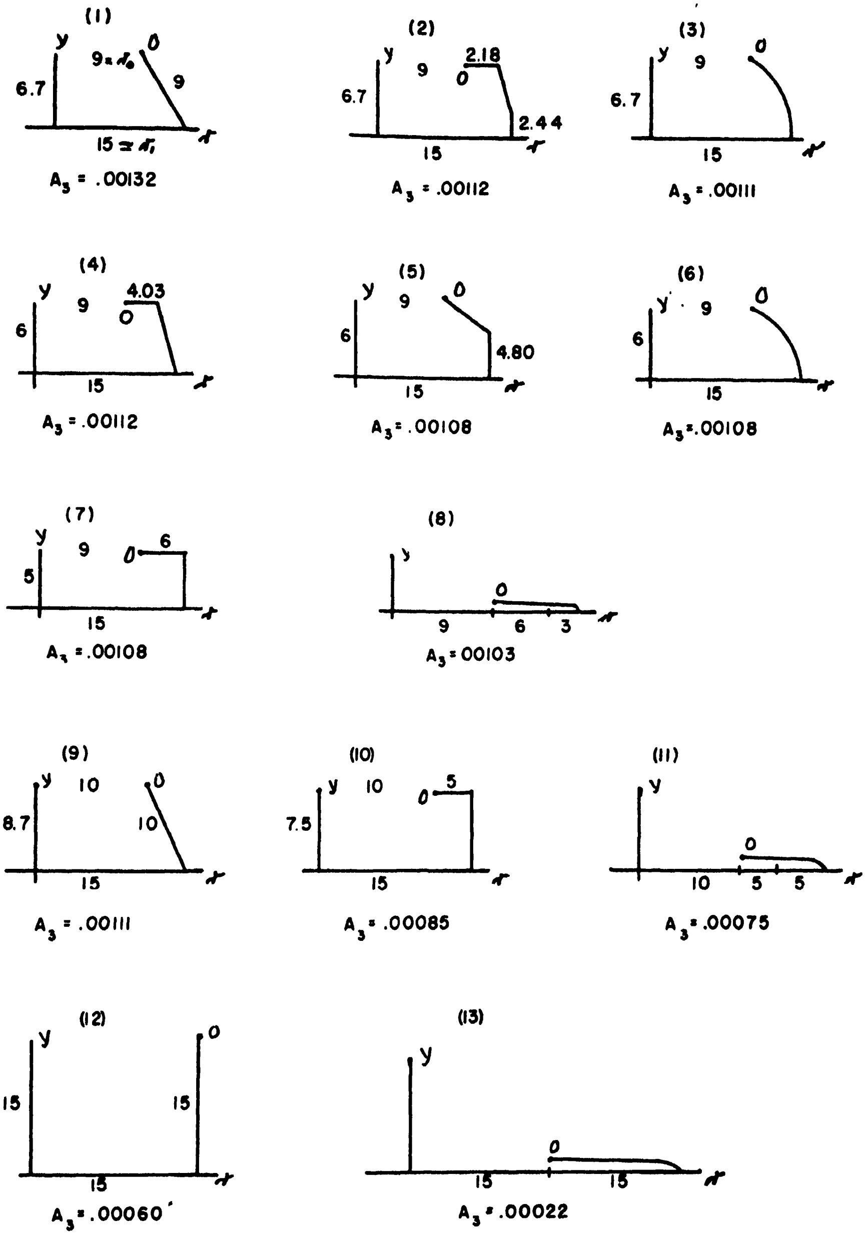


FIG. 13. (Arbitrary units.)

and minimizes

$$\left|\int_{r_0}^{r_1}\frac{(1+y'^2)^{\frac{1}{2}}dr}{r^4}\right|-\frac{1}{3r_1^3}$$

in that class of curves. The Euler-Lagrange equation for this problem is

$$\frac{d}{dr}\left[\frac{y'}{r^4(1+y'^2)^{\frac{1}{2}}}+\frac{\lambda y'}{r^2(1+y'^2)^{\frac{1}{2}}}\right]=0,$$

where λ is a constant (Lagrange multiplier). Upon integration we obtain

$$y=\left|\int_{r_0}^{r}\frac{dr}{\left(\left(\frac{1+\lambda r^2}{kr^4}\right)^2-1\right)^{\frac{1}{2}}}\right|, \qquad \text{(B)}$$

where $k$ is a constant. To determine λ, $k$ we have

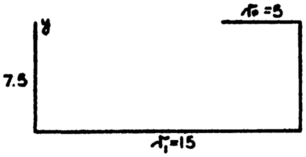


FIG. 14.

the equations

$$\left|\int_{r_0}^{r_1}\frac{dr}{r^2\left(1-\left(\frac{kr^4}{1+\lambda r^2}\right)^2\right)^{\frac{1}{2}}}\right|=\frac{1}{r_1},$$

$$\left|\int_{r_0}^{r_1}\frac{dr}{\left(\left(\frac{1+\lambda r^2}{kr^4}\right)^2-1\right)^{\frac{1}{2}}}\right|=H.$$

By approximate integration, these equations can be solved in any numerical case and the curve (B) found graphically. Figure 12 shows the curve (B) for $r_0, r_1$ fixed, with $r_0 \leqq r_1 \leqq 2r_0$, and for various $H$. As $H$ decreases toward zero, the value of $A_1$ furnished by the curve (B) decreases. Figure 13 illustrates this fact and the fact that the optics improve as $r_0$ increases. In each drawing of the table $A_1=0$. In (3), (6)–(8), (10), (11) the contour curve is of the minimizing type (B). For the cylindrical type of surface (7), (10), (12) we have

$$0=A_1=\int_{t_0}^{t_1}\frac{(r'^2+y'^2)^{\frac{1}{2}}dt}{r^2}-\frac{1}{r_1}=\int_{r_0}^{r_1}\frac{dr}{r^2}$$
$$+\int_0^H\frac{dy}{r_1^2}-\frac{1}{r_1}=\frac{1}{r_0}-\frac{2}{r_1}+\frac{H}{r_1^2},$$

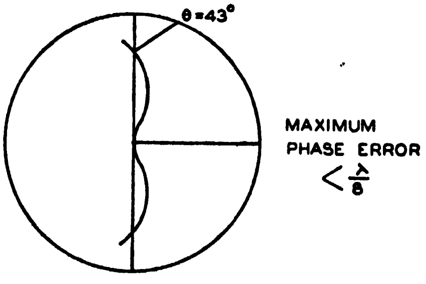


FIG. 15.

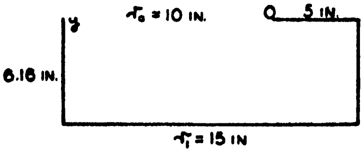


FIG. 16.

so that

$$H=r_1^2\left[\frac{2}{r_1}-\frac{1}{r_0}\right].$$

The limiting degenerate case $H=0$, $r_1=2r_0$ is illustrated by (8), (11), (13), and is a Schmidt circular reflector.

One particular application has already arisen for which $r_0$, $r_1$, $H$ were severely restricted. The wave-length to be used was $\lambda=1.25$ cm, $r_1$ was assigned as 15 inches, $r_0$ was to be $\leqq 10$ inches (for mechanical reasons). Due to lack of space no contour curve which forced the curve (B) to bulge out beyond $r_1$ was allowable, so that it was necessary that $H$ be $\geqq r_1^2(2/r_1-1/r_0)$. The cylindrical case (10) (see Fig. 14) seemed the most practicable. Computation of path lengths for this case shows that the phase errors at the aperture remain less than $\lambda/8$ for approximately $\frac{2}{5}$ of the aperture. If the height $H$ of the cylinder is lowered slightly, the effect is to destroy the equation $A_1=0$, and change to $A_1<0$ and a wave front of Type 2. The height can be adjusted ($H=6.16$ inches) so that the extremum points on the wave front are $\lambda/8$ off the aperture (see Figs. 15, 16). In this case the phase errors remain less than $\lambda/8$ for approximately $\frac{3}{5}$ of the aperture. If the allowable error were $\lambda/4$, by adjusting $H$ to 5.55 inches the phase errors would remain less than $\lambda/4$ for almost 7/10 of the aperture.

## 4. SURFACE OF REVOLUTION WITH DIELECTRIC

If dielectric is introduced into part of the region between the parallel plates of a surface of revolution scanner of the type described in the previous section (see Fig. 17), various degrees of improvement in the optics can be obtained. For

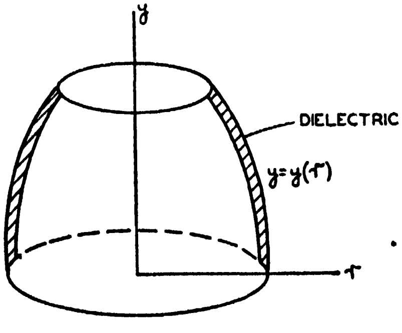


FIG. 17.

example if dielectric of index of refraction $\mu(r)$ is used as shown in Fig. 17, the result is a discontinuity in $\alpha$ at the base disk. Introducing optical length $s$ along a geodesic we obtain

$$s_1=\left|\int_{r_0}^{r_1}\frac{\mu^2 r(1+y'^2)^{\frac{1}{2}}dr}{(\mu^2r^2-h^2)^{\frac{1}{2}}}\right|,$$

$$\theta_1=\left|h\int_{r_0}^{r_1}\frac{(1+y'^2)^{\frac{1}{2}}dr}{r(\mu^2r^2-h^2)^{\frac{1}{2}}}\right|,$$

$$\alpha_1=\arcsin\frac{h}{r_1}$$

where $\alpha_1$ is the angle the geodesic makes with a radius of the base immediately *after* emerging into the base.

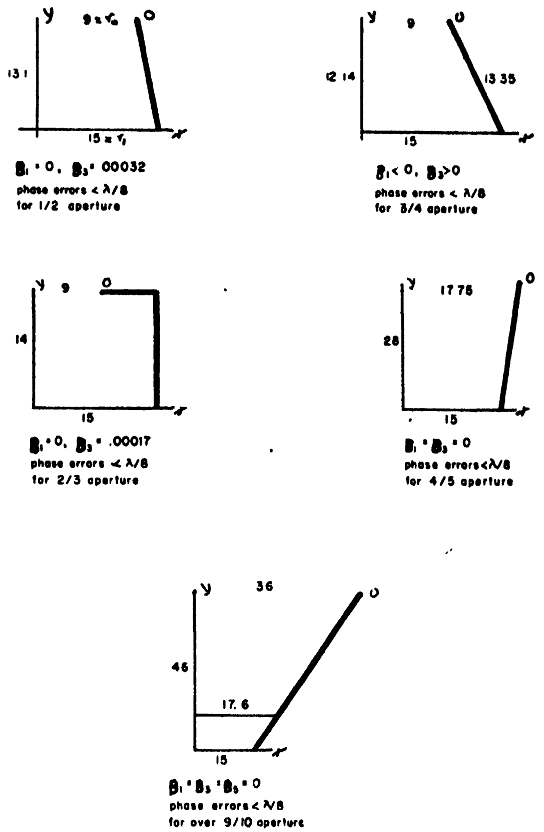
The series for $\theta_1-\alpha_1$ becomes

$$\theta_1-\alpha_1=\sum_{n=1}^{\infty}\frac{h^{2n-1}}{2n-1}B_{2n-1}\frac{1\cdot3\cdot5\cdots(2n-1)}{2^{n-1}(n-1)!(2n-1)},$$

where

$$B_{2n-1}=\left|(2n-1)\int_{r_0}^{r_1}\frac{(1+y'^2)^{\frac{1}{2}}dr}{\mu^{2n-1}r^{2n}}\right|-\frac{1}{r_1{}^{2n-1}}.$$

If $\mu=1.6$, and these results are applied to the case considered at the end of Section 3, Fig. 18 shows the type of improvement obtainable. The last two figures show that if we allow $r_0>r_1$, an extremely high degree of optical accuracy is obtainable. Of course, the feasibility of the use of dielectric depends on the amount of absorption of energy that occurs.



$B_1=0$, $B_3=.00032$
phase errors < $\lambda/8$
for 1/2 aperture

$B_1<0$, $B_3>0$
phase errors < $\lambda/8$
for 3/4 aperture

$B_1=0$, $B_3=.00017$
phase errors < $\lambda/8$
for 2/3 aperture

$B_1=B_3=0$
phase errors < $\lambda/8$
for 4/5 aperture

$B_1=B_3=B_5=0$
phase errors < $\lambda/8$
for over 9/10 aperture

FIG. 18. (Units are inches; $\lambda=1.25$ cm.)

This paper is a condensation of a laboratory report based on work done at the Radiation Laboratory, Massachusetts Institute of Technology, during the summer of 1944. The physical motivation of the problems considered, as well as some of the mathematical development, is due to L. J. Chu, C. V. Robinson, and R. C. Spencer. For obvious reasons, the emphasis is on approximate solutions suitable for practical use rather than the general solution of the mathematical problems involved.

# Space Charge Frequency Dependence of Magnetron Cavity*

W. E. LAMB, JR. AND M. PHILLIPS†
*Columbia Radiation Laboratory, Columbia University,*** *New York, New York*

(Received July 15, 1946)

The effect on the resonant frequencies of a magnetron cavity produced by the presence of a thin layer of charge surrounding the cathode is investigated theoretically. The corresponding experimental procedure is the "cold impedance testing" of a magnetron with a hot cathode in the presence of a magnetic field and a low d.c. cathode-anode voltage. A small amplitude theory is used, based on the single stream steady state. The correction to the frequency of the empty cavity due to a thin cloud of charge surrounding the cathode is given by a resonance-like formula about the "cyclotron frequency," $eB/2\pi m$. In the "Rising-Sun" tube the curve of $\nu$ *vs.* $\nu_c$ for constant d.c. voltage is asymmetric, apparently because of the presence of the "zero" mode. The validity of the theory is limited to low level oscillations, but the comparison of theory with experimental results seems to support the existence of the single stream steady state for non-operating magnetrons.

## I. INTRODUCTION

THE investigation reported here was undertaken with a view to the use of a non-operating magnetron structure as a tuning device. The tuning unit could either be built into the magnetron to be tuned, utilizing an auxiliary cathode coaxial wth the working cathode, or be coupled externally as an accessory tuner. Since any one of a great variety of coupling possibilities will introduce further complexities, however, and since the validity of the theory is limited to low level oscillations, we shall consider primarily the effect of space charge on the resonant frequencies of the magnetron cavity. The results can be applied to tuning if coupling considerations are taken into account, but in themselves describe the results of cold impedance testing a magnetron with the cathode hot, in the presence of a magnetic field, and with a d.c. voltage below that needed for operation. Our considerations also appear to throw some light on general magnetron theory. The perturbation of circulating space charge has been investigated by Blewett and Ramo[1] and Brillouin.[2] In both these papers it is assumed that the cathode radius is very small in comparison with other dimensions of the apparatus, an assumption not justified for modern magnetrons. Furthermore, their investigations are concerned only with the charge cloud itself, not its effect on anode block frequencies. It has already been pointed out[3] that in reference 1 the neglect of the effect of the constant magnetic field on the perturbations of the electron motion is not justifiable.

A small amplitude theory for magnetrons has been given by Bunemann,[3] who makes use of a velocity potential, analogous to that used in hydrodynamics. The r-f velocity potential is considered as a small perturbation of the d.c. velocity potential associated with the single stream steady state. The impedance presented at the boundary of the charge cloud is computed and, in principle, matched to that of the remainder of the cathode-anode space.

Although some question has been raised[4] as to the stability of the single stream steady state (also sometimes called the Hull or Brillouin steady state) we shall, like Bunemann, take it as our starting point. The validity of a small amplitude theory as applied to a magnetron under operating conditions need not concern us here, since we make no attempt to describe self-sustained oscillations. It turns out that there is very little further complication in handling perturbations of the fields themselves, instead of the velocity potential, while there are advantages

---

* Publication assisted by the Ernest Kempton Adams Fund for Physical Research.

† At Brooklyn College, Brooklyn, New York.

** This paper is based on work done under OSRD contract OEMsr-485.

[1] Blewett and Ramo, Phys. Rev. **57**, 635 (1940); J. App. Phys. **12**, 856 (1941).

[2] L. Brillouin, Phys. Rev. **60**, 385 (1941) and **62**, 166 (1942).

[3] O. Bunemann, C.V.D. Report, Ref. Mag. 37 (1944), unpublished.

[4] For example, F. Bloch, NDRC Div. 15: RP-295, 411-175 (May 25, 1945), unpublished.

in using the fields for the purpose of matching to the anode cavities. Also the theory of the zero mode, of interest in "rising sun" tubes, appears naturally in our formulation, while Bunemann's hydrodynamical analogy failed to show it. This was caused by his neglect of the vector potential, and setting the curl of the electric field equal to zero. Only in the discussion of boundary conditions at the cathode do we need refer to a velocity potential.

An analysis of the limitations of our theory, and the difficulties involved in satisfying the usual cathode boundary conditions, is undertaken in the concluding section of the present paper.

## II. GENERAL METHOD

We shall use a fixed set of coordinates, $r$, $\theta$, $z$, and assume that the tube is sufficiently long that the dependence on $z$ can be neglected. The unperturbed fields, charge density, and velocity of the single stream steady state are, in rationalized m.k.s. units:[5]

$$\mathbf{E}_0=(m/e)(\omega_c^2/4)(1-b^4/r^4)\mathbf{r}, \tag{1}$$

$$\rho_0=(m/e)(\epsilon_0/2)\omega_c^2(1+b^4/r^4), \tag{2}$$

$$\mathbf{v}_0=-\omega_c/2(1-b^2/r^2)r\hat{\theta}, \tag{3}$$

where $\omega_c=eB_0/m$, the "cyclotron" angular velocity, $b$ is the cathode radius, and $B_0$ the externally applied magnetic field in the $z$ direction. The equation of motion for an electron at a fixed point in space, correct to the first order in the perturbation $(m/e)d\mathbf{v}/dt=\mathbf{E}+\mathbf{v}\times\mathbf{B}$, becomes

$$(m/e)d\mathbf{v}_0/dt+(m/e)d\mathbf{v}_1/dt$$
$$=\mathbf{E}_0+\mathbf{E}_1+\mathbf{v}_0\times\mathbf{B}_0+\mathbf{v}_1\times\mathbf{B}_0+\mathbf{v}_0\times\mathbf{B}_1. \tag{4}$$

The single stream steady state satisfies the zero-order equation, so that the perturbation equation becomes

$$\partial v_1/\partial t+(\mathbf{v}_0\cdot\nabla)\mathbf{v}_1+(\mathbf{v}_1\cdot\nabla)\mathbf{v}_0$$
$$=(\mathbf{E}_1+\mathbf{v}_1\times\mathbf{B}_0+\mathbf{v}_0\times\mathbf{B}_1)(e/m). \tag{5}$$

The first-order field equations are

$$\mathrm{curl}\mathbf{E}_1=-\mu_0\dot{H}_1, \tag{6}$$

$$\mathrm{curl}H_1=\epsilon_0\dot{E}_1+\rho_0\mathbf{v}_1+\rho_1\mathbf{v}_0, \tag{7}$$

where $\rho_1=\mathrm{div}\mathbf{E}_1$. It will be noted that the zero-order quantities do not satisfy the zero-order curl$H$ equation; but since $\rho_0\mathbf{v}_0$ gives a correction to the applied magnetic field which is of order $v_0^2/c^2$, the neglect of this term is consistent with a non-relativistic calculation.

If the time and angular dependence of all quantities is given by the factor $\exp i(n\theta-\omega t)$, where $n$ is the mode number, the field equations in component form become

$$i\omega\mu_0H=(1/r)(\partial/\partial r)(rE_\theta)-inE_r/r, \tag{8}$$

$$inH/r=-i\omega\epsilon_0E_\theta+\rho_0v_r, \tag{9}$$

$$-\partial H/\partial r=-i\omega\epsilon_0E$$
$$+\epsilon_0v_0/r\{\partial(rE_r)/\partial r+inE_\theta\}+\rho_0v_\theta. \tag{10}$$

The equations of motion take the form

$$-i\omega v_r+(v_0/r)inv_r=(e/m)E_r+v_\theta\omega_c+(e/m)v_0\mu_0H, \tag{11}$$

$$-i\omega v_\theta+(v_0/r)inv_\theta+v_r\partial v_0/\partial r=(e/m)E_\theta-v_r\omega_c. \tag{12}$$

Here the subscripts denoting perturbation fields and velocities have been dropped, only the zero subscripts being retained.

Equations (11) and (12) may be solved for $v_r$ and $v_\theta$ in terms of the fields, and substituted in the field equations. These may then be combined to give a single second-order differential equation in $E_\theta$, or $E_r$. This equation is too complicated to be very useful without extended numerical integration, but for the limiting case of thin layers of charge on the cathode conclusions as to the effect on the resonant frequencies may be easily drawn.

The boundary condition at the cathode demands that the tangential component of the electric field be equal to zero, and the differential equation for $E_\theta$ admits of a solution in powers of $(r-b)=y$, starting with the first power. For a very thin charge cloud, $E_\theta$ varies as $y$. The first-order $y$ dependence of the other field quantities can then be found by making use of the field equations. This approximation breaks down for frequencies too near the cyclotron frequency, since the coefficients of higher powers of $y$ contain the factor $(\omega_c^2-\omega^2)$ in the denominator.

To justify the use of this approximation we may investigate the radius of the charge cloud under steady conditions. One has only to demand

---

[5] Reference 2, part I, sections 2 and 3.

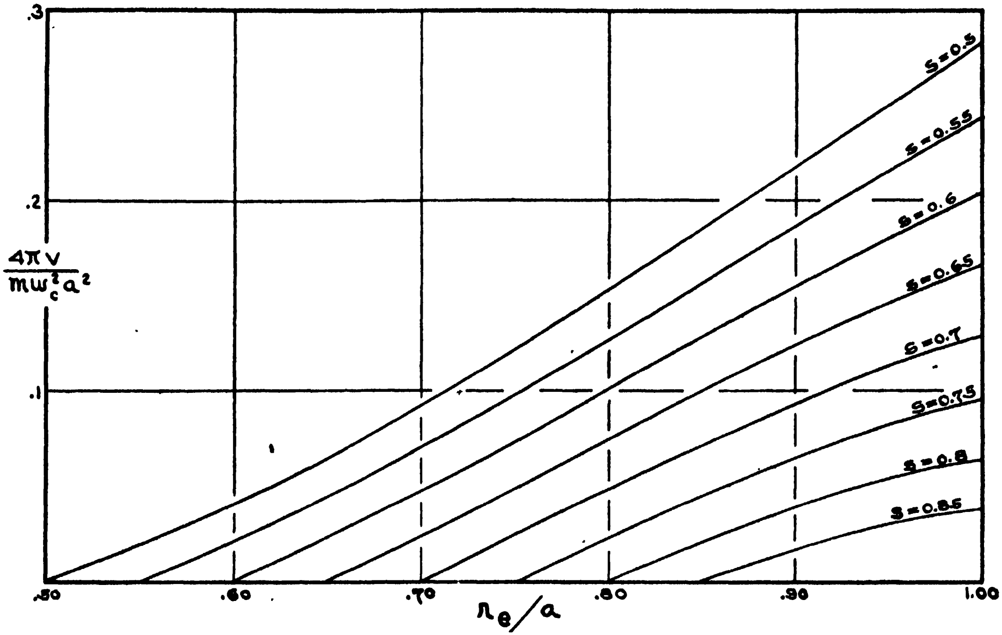


FIG. 1. A plot showing the radius of the boundary of the single stream steady state charge cloud as a function of applied cathode-anode voltage, for cathode diameter to anode diameter ratios ranging from 0.5 to 0.85.

that the fields be continuous across the boundary of the cloud, at $r_e$, and that the total difference of potential between cathode and anode is the applied voltage. The equation for the applied voltage is

$$V=(m/e)(\omega_c^2/8)r_e^2[(1-b^2/r_e^2)^2 +2(1-b^4/r_e^4)\ln(a/r_e)]. \quad (13)$$

Here $a$ is the inside radius of the anode. A plot of $4eV/m\omega_c^2a^2$ against $r_e/a$ has been made for a variety of $b/a$ ratios (Fig. 1), but for low voltages the relation between $y_e$ and $V$ is linear, and well represented by the simple formula

$$y_e/b=(e/m)(V/\omega_c^2b^2)/\ln(a/b).$$

For $\omega_c$ of the order of $10^4$ megacycles, as it is for the experimental tests made, $y_e/b$ is seen to be extremely small, of the order of $10^{-3}$.

### III. APPLICATION TO MODES WITH $n\neq 0$

If $n$ is not equal to zero it is justifiable to neglect the last term of Eq. (11), and to assume that the curl of **E** equals zero. This corresponds to a similar simplification in Bunemann's theory. Even under the simplifying assumptions the exact equation for $E_\theta$ is very cumbersome, although it can be derived in a straightforward way as indicated above. For large $nb/y$ it becomes fairly simple, especially if written in terms of the variable $\eta$, defined by $r/b=\exp(\eta/n)$:

$$(\omega'^2-\omega_c^2)d^2(rE_\theta)/d\eta^2+2\omega_c^3/\omega' d(rE_\theta)/d\eta = [2\omega_c^4/\omega'^2+(\omega'^2-\omega_c^2)](rE_\theta). \quad (14)$$

Here

$$\omega'=\omega+n\omega_c(1-b^2/r^2)/2=\omega+\eta\omega_c.$$

All terms of first order in $\eta$ are kept, in Eq. (13), but not in $\eta/n$, i.e., the effects of curvature alone are neglected, an approximation justified for fairly large $n$. Away from $\omega_c$, $y=0$ or $\eta=0$ is not a singular point, and the power series solutions start with $y$ or a constant. This is also true of the more complicated exact equation, awkward as it is to write down. For the plane magnetron, with a time and $x$ dependence given by $\exp i(\kappa x-\omega t)$, the corresponding equation is

$$(\omega\omega'-\omega_c^2)d^2E_x/dy^2+2\kappa\omega_c^3/\omega' dE_x/dy = (2\omega_c^4/\omega'^2+(\omega\omega'-\omega_c^2))\kappa^2E_x, \quad (14a)$$

where $\omega'=\omega+y\kappa\omega_c$.

The impedance[6] presented by the charge cloud at its boundary, correct to the first order in the

---

[6] See, for example, S. A. Schelkunoff, Bell Sys. Tech. J. 17, 17 (1938), or J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, Book Company, Inc., New York, 1941), p. 282.

thickness of the charge layer outside the cathode, is then easily computed directly from Eqs. (8) to (12) above. If $E_\theta = \mathcal{E}y/b$, $\mathcal{E}$ being inserted as a constant to supply the right dimensions, then from (8) $E_r = \mathcal{E}/in$

from (12)

$$v_\theta = (e/m)iE_\theta/\omega = (e/m)iy\mathcal{E}/\omega b,$$

from (11)

$$v_r = \frac{i}{\omega}(e/m)E_r + \frac{i\omega_c}{\omega}v_\theta,$$

and from (9)

$$H = (\mathcal{E}/in)(i\omega E_r + \omega_c^2 m v_r/e)$$

$$= b\epsilon_0\mathcal{E}(\omega_c^2 - \omega^2)/in^2\omega.$$

$$\therefore \quad Z_{\mathrm{el}} = E_\theta/H = \frac{in^2y}{b^2\epsilon_0\omega}\,\frac{\omega^2}{\omega_c^2-\omega^2}$$

$$= i(\mu_0/\epsilon_0)^{\frac{1}{2}}\frac{n^2}{bk}\,\frac{y}{b}\,\frac{\omega^2}{\omega_c^2-\omega^2}\ \text{ohms}, \qquad (15)$$

where

$$k = \frac{2\pi}{\lambda} = \frac{\omega}{c}.$$

Since it is approximately $ny\omega^2/b(\omega_c^2-\omega^2)$ which enters as an expansion parameter, this expression may be considered valid for $ny/b < (\omega_c^2-\omega^2)/\omega^2$. The failure of $E_r$ to satisfy the conventional condition at the cathode corresponding to space charge limited emission will be discussed later.

To determine the resonant frequencies of the magnetron when the cathode is surrounded by the sheath of space charge this impedance is used as a boundary condition instead of the simple demand that $E_\theta = 0$ in the usual method.[7, 8]

We shall confine our attention to the symmetric anode. The equation for matching the fields at the openings to the anode cavities[9] for the $n$th mode is:

$$\cot kL = \frac{N\Theta}{\pi}\sum_{m=n+lN}^{\infty}\frac{J_m'(ka)+\alpha_m N_m'(ka)}{J_m(ka)+\alpha_m N_m(ka)} \times\left(\frac{\sin m\Theta}{m\Theta}\right)^2, \qquad (16)$$

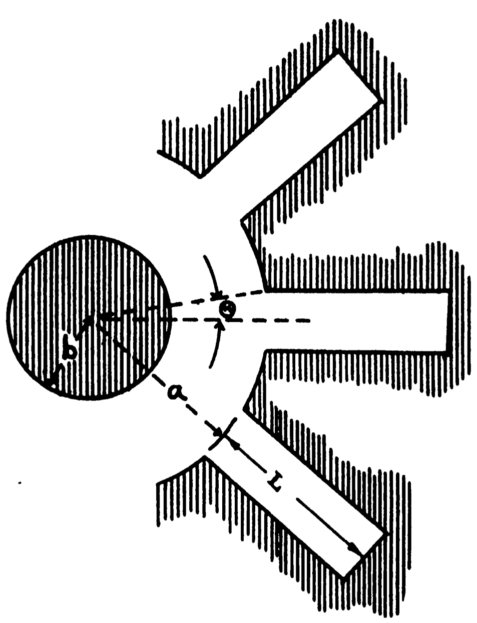


FIG. 2. Partial section of symmetrical anode magnetron. $b$ is the cathode radius, $a$ the inside radius of the anode, $L$ the length of a slot, and $\Theta$ is half the angle subtended by each slot.

where $2\Theta$ is the angular aperture of the cavities, $a$ is the anode radius, and $L$ is the length of the anode slots, as shown in Fig. 2. $\alpha_m$ is the parameter arising from the presence of the cathode or the cathode charge layer boundary condition. It has been shown that a consideration of the cathode radius in an empty cavity introduces very little change in the frequency except for the lowest mode numbers, and that for any $n$ the preponderance of the effect arises from the fundamental. This should remain true for the effect of thin layers of charge on the cathode. For this reason we may neglect all harmonics and determine $\alpha_n$ by simply matching the electron impedance at the edge of the cloud to that given by the solution for a cylindrical cavity. Furthermore, since $kb \ll 1$ for most tubes we may substitute for each Bessel and Neumann function the first term in its power series expansion about the origin.

The matching then involves using

$$H \simeq J_n(kr) + \alpha_n N_n(kr),$$

$$E \simeq -i\mu_0\omega/k[J_n'(kr) + \alpha_n N_n'(kr)],$$

so that

$$Z_{\mathrm{cavity}} = E_\theta/H = \frac{i\sqrt{\mu_0}}{\sqrt{\epsilon_0}}\,\frac{J_n'(kr)+\alpha_n N_n'(kr)}{J_n(kr)+\alpha_n N_n(kr)} \qquad (17)$$

---

[7] H. Goldstein, "Resonant modes of magentron cavities," NDRC Div. 14, Report No. 493 (Dec. 14, 1943).

[8] N. Kroll and W. E. Lamb, Jr., "Resonant modes of the "rising sun" anode," NDRC Div. 14, Report No. 323 (Oct. 25, 1944).

[9] Cf. Goldstein, reference 7, p. 9.

is determined by the condition $Z_{01}=Z_{\text{cavity}}$ at the boundary of the space charge, i.e.,

$$\frac{n^2 y_e}{kb^2}\frac{\omega^2}{\omega_c^2-\omega^2}=-\frac{J_n'(kr_e)+\alpha_n N_n'(kr_e)}{J_n(kr_e)+\alpha_n N_n(kr_e)}, \quad (18)$$

where $r_e=b-y_e$ is the radius of the charge cloud. It follows that

$$\alpha_n=-\frac{J_n'(kr_e)}{N_n'(kr_e)}\left\{1+\frac{2ny_e}{b}\frac{\omega^2}{\omega_c^2-\omega^2}\right\}$$
$$=\left(\frac{kb}{2}\right)^{2n}\frac{\pi}{n!(n-1)!}\left\{1+\frac{2ny_e}{b}\frac{\omega^2}{\omega_c^2-\omega^2}\right\}. \quad (19)$$

This expression for the boundary condition parameter is to be compared with the corresponding expression in the absence of the layer of space charge,

$$\alpha_n=-\frac{J_n'(kb)}{N_n'(kb)}=\left(\frac{kb}{2}\right)^{2n}\frac{\pi}{n!(n-1)!}. \quad (20)$$

The matching equation at the anode radius, from which the resonant frequencies of the whole resonating cavity are determined, becomes

$$\cot kL=\frac{N\Theta}{\pi}\sum_{m=n+lN}^{\infty}\left(\frac{\sin m\Theta}{m\Theta}\right)^2\frac{J_m'(ka)}{J_m(ka)}$$
$$\times\left\{1+2\left(\frac{b}{a}\right)^{2m}\left(1+\frac{2my_e}{b}\frac{\omega_c^2}{\omega_c^2-\omega^2}\right)\right\}. \quad (21)$$

This is to be compared with Eq. (15), of reference 4, *viz.*,

$$\cot kL=\frac{N\Theta}{\pi}\sum_{m=n+lN}^{\infty}\left(\frac{\sin m\Theta}{m\Theta}\right)^2\frac{J_m'(ka)}{J_m(ka)}$$
$$\times\left\{1+2\left(\frac{b}{a}\right)^{2m}\right\}. \quad (22)$$

From the fact that $(ny_e/b)(\omega^2/(\omega_c^2-\omega^2))$ appears as a small quantity in the solution of our differential equation for $E_\theta$ it is not permissible to use this correction for $\omega_c=\omega$, but if $y_e$ is sufficiently small (21) may be valid for frequencies very near the cyclotron frequency. It is seen that the maximum tuning predicted safely by Eq. (21) will be of the order of the correction to the frequency caused by the finite cathode.

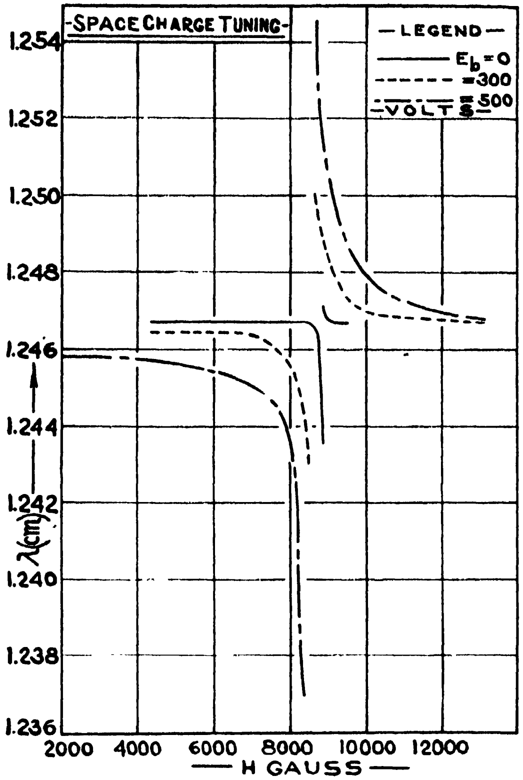


FIG. 3. Wave-length of rising-sun magnetron cavity as a function of magnetic field for cathode-anode voltages shown in the legend.

For $\omega_c<\omega$, i.e., for low magnetic fields, it will be in the direction of cancelling the cathode effect, decreasing the resulting resonant wave-length. For modes of higher $n$, where the fields at the center are small, there will be no appreciable effect; none, at least, which can be predicted with confidence from Eq. (21).

Some preliminary measurement of the tuning effects of the space charge layer about a magnetron cathode were made by Mr. B. Hildebrand of this laboratory. A number of 1.25-cm wave-length magnetrons were "cold impedanced" with a hot cathode in the presence of an axial magnetic field and d.c. radial electric field. The resonant wave-length was measured as a function of the applied fields. The r-f used was obtained from a local oscillator and had an intensity of a few milliwatts. The voltages used were insufficient to allow appreciable leakage current and back bombardment of the cathode, except at the lower magnetic fields. No detectable r-f of microwatt intensity was generated by the magnetrons, which were designed for pulsed operation at well over 10 kv.

Figure 3 shows the resonant wave-length ($\pi$ mode) of a rising sun magnetron cavity as a function of magnetic field for three values of d.c. voltage, 0, 300, and 500 volts. Within the limits of the field calibration, the sharpest tuning effects occur as expected at the cyclotron frequency. The tuning observed for zero anode voltage may be ascribed to the thermal motion of the electrons, for no tuning was observed with five volts of reversed potential.

These experiments are being continued with better cavity geometry and more careful control of cathode emission. A quantitative comparison between theory and experiment will be made when more complete data are available.

If one tries to make more rigorous use of the equation for $E_\theta$ in determining the impedance of the charge cloud, the first correction is a displacement of the singularity toward higher frequencies by an amount of the order of magnitude of $ny_e\omega/b$. The singularity in terms of $B_0$ is shifted toward lower magnetic fields according to the equation

$$\omega_c'=\omega_c\left(1+\frac{3}{2}\frac{ny_e}{b}\right).$$

Experimentally this means that the singularity in the curves of $\omega$ against $B_0$ should be displaced slightly as the d.c. potential and thus the thickness of the charge cloud is increased. The experiment is not conclusive on this point.

## IV. THE ZERO MODE

The case of the zero mode, of importance in rising sun tubes,[8] is somewhat different. This mode corresponds to a circular current in the tube. Here curl$E$ is not equal to zero, and the correction term in the magnetic field, $v_0B_1$, must also be kept. The absence of any dependence on $\theta$ makes the equations simpler, however, and if the layer of charge is sufficiently thin to justify the assumption that the magnetron is plane the differential equation for $E_\theta$ becomes

$$[1-k_c^2y^2/(1-\gamma^2)]d^2E_\theta/dy^2-\frac{2k_c^2y}{1-\gamma^2}dE_\theta/dy$$
$$-k_c^2(1-\gamma^2)E_\theta=0, \quad (23)$$

where $k_c=\omega_c/c$, $\gamma=\omega/\omega_c$.

This equation somewhat resembles Legendre's equation, and is satisfied by a power series containing only odd powers of $y$, but again the solution is not valid for $\omega=\omega_c$, since the coefficient of $y^3$ is $k_c^2(2+(1-\gamma^2)^2)/6(1-\gamma^2)$. Also,

$$E_r=\frac{i[E_\theta+ydE_\theta/dy]}{\gamma(\gamma^2-1)}. \quad (24)$$

To obtain a charge cloud impedance which does not depend critically on the frequency, it is necessary to take account consistently of the second term in the expansion in $y$. Working to this order also involves taking into account the deformation of the boundary layer. If a point on the surface of the charge cloud differs from its unperturbed value by an amount $\delta$, the value of $H$ to be used in matching differs from that inside the cloud by the contribution of the surface current, $\rho_0v_0\delta$.

To compute the impedance presented at the edge of the cloud we need

$$E_\theta\sim y+\frac{k_c^2y^3}{3(1-\gamma^2)},$$

$$H=\frac{1}{i\mu_0\omega}\frac{dE_\theta}{dy}-\rho_0v_0\delta.$$

Since $r_e=\text{constant}+\delta$, it follows that $\delta=iv_r/\omega$, evaluated at $r=r_e$. Since $v_r=i\omega\epsilon_0E_r/\rho_0$,

$$H=1/i\mu_0\omega[1-k_c^2y^2/(1-\gamma^2)] \quad (25)$$

and

$$Z_{c1}=i\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}ky[1+2k_c^2y^2/(1-\gamma^2)]. \quad (26)$$

The coefficient of the resonance term would be affected by including corrections to be caused by the current and the relativistic change of mass with velocity, and so is not to be interpreted too literally. This impedance is to be matched to that of the cylindrical part of the cavity, for which the dominant term is

$$Z_{cav}=i\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\frac{J_1(kr)+N_1(kr)}{J_0(kr)+N_0(kr)}. \quad (27)$$

$$\therefore\quad Z_{cav}=\pi\left(\frac{kr_e}{2}\right)^2\left[1-\frac{2y}{r_e}\left(1+\frac{2}{3}\frac{k_c^2y^2}{1-\gamma^2}\right)\right]. \quad (28)$$

The zero-order Bessel and Neumann functions need not be expanded if use is made of the relation $J_0(x)N_1(x)-N_0(x)J_1(x)=-2/\pi x$, so that the approximations are as valid as those for modes with $n$ not equal to zero.

Matching at the anode radius leads to the equation

$$\cot kL=-\frac{N\Theta}{\pi}\frac{J_0(ka)}{J_1(ka)}$$

$$\times\left\{1+\left(\frac{b}{a}\right)^2\left[1-\frac{2y}{b}\left(1+\frac{2}{3}\frac{k_c{}^2y^2}{1-\gamma^2}\right)\right]\right]\right\}. \quad (29)$$

The effect of the charge cloud is thus always such as to cancel the cathode correction, by an amount which is larger for $\gamma<1$ than for $\gamma>1$.

It will be noticed that all amplitudes of vibration of the charge go to zero at the cathode, so that there is here no difficulty about satisfying the conventional boundary conditions. From the curl$H$ equation,

$$\frac{1}{r}\frac{\partial H}{\partial\theta}=0=\epsilon_0E_r+\rho v_r,$$

it appears that insofar as it is legitimate to neglect any $z$ dependence there is no emission at any time directly into this mode. Nevertheless, where the mode exists, by virtue of its mixture with the $\pi$ mode of rising sun tubes, for example, the tuning effect owing to its presence will be important for the same reason that a consideration of the cathode size is important, because the fields are large near the center of the cylindrical space. This effect probably accounts for the lack of symmetry in the frequency curves found in the cold test experiments carried out with rising sun tubes. Apart from the critical frequency dependence there is a correction in the presence of zero mode excitation which varies linearly with the depth of the charge cloud, and always tends to cancel the effect of the finite radius.

## V. CATHODE CONDITIONS AND VALIDITY OF THE THEORY

The linear equations used in the small amplitude theory preclude the imposing of the usual boundary conditions at the cathode if the mode number is different from zero. The two-power series solutions of the differential equation for any field quantity begin with the zero and first powers of $y$, respectively, and the field equations lead to the conclusion that if $E_\theta$ varies as the first power of $y$, then $E_r$ has a term independent of $y$. Thus if one demands that $E_\theta$ and $v_\theta$ are zero at $r=b$, one cannot also demand that $E_r$ and $v_r$ are zero at that radius, the usual condition for space charge limited emission. Taken seriously, this would mean that the radial field at the cathode is alternatively positive and negative, with the same $\exp\{i(\omega t-n\theta)\}$ dependence as other perturbation quantities. In other words the cathode is here defined merely as a surface on which the tangential field is zero.

Quite apart from a consideration of cathode conditions the small amplitude theory might not be expected to hold all the way into the cathode. In the impedance matching at the edge of the cloud we were interested only in the ratio of the perturbation fields, not their absolute magnitudes, but it might be expected that the theory would break down for $E_r$ of the order of $E_\theta$. The assumption that $|E_r|<E_\theta$ would mean, in terms of distances, that $\delta$, the amplitude of radial oscillations of the electrons is subject to the condition $\delta<y$, and that the small amplitude theory would not be valid for distances less than $y_1\sim\delta$.

The reasons for demanding the strict observance of space charge limitation of any possible emission in the absence of equilibrium conditions are not entirely clear, and in our own case there is no total current, and hence possibly no emission. Nevertheless, in view of the large part played by conventional cathode conditions in magnetron theory, it was thought feasible to attempt to match solutions of the exact equations to the fields predicted by the small amplitude theory at some distance $y_1<y_c$.

The hydrodynamical formulation has the advantage of added elegance and formal simplicity over the expansion of the field quantities and velocities separately, so that we choose this method of presentation similar to that used by Cherry.[10] Fixed coordinates may be used if the vector potential is included, which also enables us to include in the description the motion corresponding to the zero mode of the cavity.

---

[10] General Theory of the Magnetron, Council for Sci. and Ind. Research, Rept. MUM1, Sydney (April 1, 1943).

Let $m\mathbf{v}+e\mathbf{A}=-\mathrm{grad}\psi$. The proof that curl $(m\mathbf{v}+e\mathbf{A})=0$ is maintained (if there is no $H_n$) has been given, e.g., by Gabor.[11]

If $\mathbf{B}=\mathrm{curl}\mathbf{A}$ and $\mathbf{E}=-\mathrm{grad}\phi-\dot{A}$, the equation of motion leads to

$$e\phi=-mv^2/2+\psi=-\frac{(\mathrm{grad}\psi+e\mathbf{A})^2}{2m}+\psi. \qquad (30)$$

The fields, velocities, and charge density may be written in terms of $\psi$, the hydrodynamical potential or "stream function," and $\mathbf{A}$. We then demand that $E_{\mathrm{tan}}$ and $\mathbf{v}$ be zero on the cathode, and that $B_{\mathrm{tan}}$ and $\rho\mathbf{v}$ be finite. For an infinitely long cathode there is no normal component of $B$. These conditions mean that if there is any emission the charge density is infinite at the cathode surface, so that this surface is a "front" in the language of hydrodynamics.

If $\psi$ and $\mathbf{A}$ are now expanded in powers of $y$, with coefficients subject to the above condition, there follow limitations on the lowest allowed power of $y$, as well as relations between the coefficients of successive powers.

For the formal relations which must be borne in mind it is convenient to set up a system of right-handed cartesian coordinates, with $x$ in the direction of $-\hat{\theta}$. In addition to (30) and the definition of $\psi$ these relations are:

$$B=\partial A_y/\partial x-\partial A_x/\partial y=\mathrm{curl}\mathbf{A}, \qquad (a)$$

$$E_x=-\frac{\partial\phi}{\partial x}-A_x, \qquad (b)$$

$$E_y=-\frac{\partial\phi}{\partial y}-A_y, \qquad (c)$$

$$\epsilon\rho=\mathrm{div}\mathbf{E}=-\frac{\partial^2\phi}{\partial x^2}-\frac{\partial^2\phi}{\partial y^2}-\frac{\partial A_y}{\partial y}-\frac{\partial A_x}{\partial x}, \qquad (d)$$

and

$$\mathrm{div}\mathbf{A}=-\mu_0\epsilon_0\phi=\frac{\partial A_x}{\partial x}+\frac{\partial A_y}{\partial y}, \qquad (e)$$

$$\epsilon_0\dot{\rho}+\mathrm{div}\rho\mathbf{v}=0. \qquad (f)$$

The Lorentz condition (e) shows that there are only two, not three, independent series, which we can take as $\psi$ and $A_x$. Consideration of (a), remembering that $B$ must be finite but not identically zero on the cathode, leads to

$$A_x=-[B_0+\alpha_x(\theta,t)]y+\text{higher terms in } y.$$

If $\psi=\psi_l(\theta,t)y^l+$higher terms in $y$, the vanishing of $v_y$ on the cathode means that $l>1$, but $l=2$ leads to $v\sim y$, $\rho=$const., and no emission. To obtain emission the introduction of fractional powers is necessary. The conditions are satisfied by $l=5/3$, so that

$$\psi=\frac{3}{5}\psi_5(\theta,t)y^{5/3}+\frac{3}{6}\psi_6(\theta,t)y^{6/3}+\frac{3}{7}\psi_7(\theta,t)y^{7/3}+\cdots$$

is a convenient form for the series. It follows that

$$\rho v_r=\frac{\epsilon_0}{m^2e}\frac{2}{9}\psi_5^3(\theta,t)+0(y^{1/3})+\cdots. \qquad (31)$$

$\psi$ is then determined by the emission, and $A_x$ by the value of $B$ on the cathode, arbitrarily. As Cherry has pointed out, $E_y\sim y^{1/3}$, $E_\theta\sim y$, $v_r\sim y^{2/3}$, $v_\theta\sim y$, $\rho\sim y^{-2/3}$, for small $y$.

The apparent arbitrariness of the emission, determining $\psi_5$, is closely related to the question of convergence, a fact which has not been pointed out by previous authors. The coefficients $\psi_6$, $\psi_7$, etc., may be determined by substituting in the continuity equation and collecting terms in the same power of $y$. It turns out that $\psi_6\sim\psi_5/\psi_5$ and that $\psi_7$, $\psi_8$, etc., also go to infinity as $\psi_5$ goes to zero. Thus the series fails if the emission is ever equal to zero, or changes sign. This seems to indicate that the standard conditions for space charge emission, derived historically for steady constant currents, are not appropriate to the discussion of oscillations of emission about zero. This may correspond to Brillouin's finding a double stream, and not a single stream, in the case of final current.[12] An exact solution, involving emission, contradicts the single stream assumption for the cathode layer. So long as the oscillations are small, however, Brillouin found no indication that the double stream produced at the cathode extends far into the charge cloud.

An attempt was made by us to work out the single stream hydrodynamical theory on the strict basis of no emission under saturation conditions. This is formally easy: the stream func-

---

[11] Gabor, Proc. I.R.E. 33, 792 (1945).

[12] App. Math. Panel 129.3R, p. 30 *et seq.* and p. 51.

tion $\psi$ is then zero, and all quantities can be expressed in terms of **A**. $E_\theta$ and $v_\theta$ vary as the first power of $y$, as before, but $E_r$ and $v_r$ now also vary as $y$ for small $y$, with a finite charge density at the cathode. No Fourier analysis in the time is possible, however, so that the solutions cannot be matched to those of the linear equation.

The question of the stability of the single stream steady state under a periodic perturbation thus involved the cathode conditions very intimately. Those proofs of instability which are based on the assumption of ideal equilibrium cathode conditions seem to us of doubtful validity. In the absence of knowledge of the nature of the cathode surface, and the applicability of the demand that $E_r=0$, the qualitative experimental verification of theoretical results whose existence depends on there being such a steady state seems as good a way as any of testing the matter. It would remain to show that our results do indeed depend on the existence of a single stream steady state.

While it is impossible to make a general and categorical answer to this question, the observed resonance curves indicate a resonance frequency dependence which does depend on the existence of a magnetic field and the attendant circulation of the charge, and not merely on the charge density. This is easily shown. Maxwell's equation can be set up for a charge cloud of cylindrical symmetry, having the density $\rho=\epsilon_0\omega_c^2m/e$. For a thin sheath of static charge about the cathode the resulting equations can be solved, and the impedance looking into the sheath can be computed. For this case

$$Z_{el}=-i\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\left\{\frac{n^2y}{b^2}\frac{c}{\omega}\left(\frac{\omega^2}{\omega^2-\omega_c^2}\right)^{\frac{1}{2}}\right\}, \tag{32}$$

corresponding to the results in the plane case well known in the theory of the ionosphere.[6] It is seen that at $\omega=\omega_c$ the impedance matching breaks down, and there is a change from reactive to resistive impedance. Contrary to this prediction the experimental curve, Fig. 3, shows a typical resonance pattern.

If, on the other hand, the effect of the applied magnetic field on the motion is taken into account, but not the zero-order velocity, the field equations and the equations of motion lead to

$$Z_{el}=i\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\frac{n^2y}{b^2}\frac{c}{\omega}\frac{\omega^2-\omega_c^2}{\omega^2-2\omega_c^2}. \tag{33}$$

This is again ruled out by the experimental results. Thus the experiments indicate a perturbation of circulating individual electrons, and constitutes evidence for something like the single stream steady state.

Another solution of the stationary state problem has been given recently by Gabor,[13] who introduces "ergodic disorder," retaining conservation of energy and angular momentum for each individual electron. This is in reality an infinite stream steady state. It is interesting to see the differences and similarities between this solution and the one we have used. Gabor's method depends on the existence of a non-zero velocity of emission, although the assumption of equilibrium conditions means zero field at the cathode. If the initial velocity $u_0$ is small the tangential velocity at any point in the swarm of electrons is distributed about the $v_0$ of the single stream steady state, the width of the distribution being just $2au_0/r$, while there are small random velocities in the $r$ and $z$ directions to preserve the energy intact. The value of the saturation density arrived at by Gabor[14] also agrees with that of the single stream steady state for a thin layer about the cathode, although this is not true for thick clouds. Unfortunately the kinetic energy of an electron in a thin cloud is not in general much greater than the thermal energies corresponding to the cathode temperature, so that this agreement must not be taken too seriously. The same kind of difficulties as to how such a state is set up occur here as with the single stream steady state.

---

[13] D. Gabor, Proc. Roy. Soc. **183**, 357 (1945).

[14] Reference 13, Eq. (28.1).

# The Reduction of Microphonics in Triodes

A. H. WAYNICK
*Ordnance Research Laboratory, School of Engineering, The Pennsylvania State College, State College, Pennsylvania*

(Received September 6, 1946)

An analysis is presented showing the effect of grid displacement on the plate current of a planar triode. This is defined in terms of a new parameter called the motional transconductance. The analysis indicates that a useful mechanical displacement indicator may be constructed by moving the grid and noting the change in plate current. The analysis also indicates that the effect of certain modes of grid displacement on the plate current may be eliminated under conditions which are stipulated. Since microphonic effects are, in great part, caused by the mechanical motion of grid structures, the analysis indicates how microphonism may be reduced. It is shown that the electrical transconductance is large under these conditions, for the case investigated, so that the tube functions normally as an amplifier. Preliminary experimental results for comparison with the theory are presented.

## INTRODUCTION

A RECENT patent[1] describes an electron tube whose plate current is varied by displacement of the grid, relative to the rest of the tube structure, while holding the applied tube potentials at constant values. The grid is mechanically connected through a vacuum-tight member so that it may be moved by the application of suitable forces to the tube envelope.

The following analysis was initially carried out to determine the feasibility of this idea when it had been determined that this type of tube had not been reduced to practice. The patent disclosure itself did not prove this.

The analysis indicates the plate current variations that may be expected in terms of a parameter called the motional transconductance. Inspection of the relation giving this parameter shows that, under suitable conditions, the effect of grid displacement on the plate current vanishes. This suggests a method of design, or mode of operation, by which the microphonism of a tube may be reduced since, as is well known, most of the microphonic difficulties encountered in tubes are caused by the mechanical movement of the grid. Design and operation relations are developed for this case.

Preliminary experiments are then described which indicate that the theory is fundamentally correct.

## THEORY OF A PLANAR TRIODE[2]

Consider a planar triode as illustrated in Fig. 1. Here $K$ is the cathode; $A$ is the anode; and $G$ is the grid structure. The grid is assumed to be made of equi-distant parallel rods having a radius $R$ and a rod center spacing of $1/n$.

The distance $dg$ is the cathode-grid center distance. The distance $dp$ is the cathode to plate distance. The distance $dq$ is the cathode to an equivalent plate distance, of an equivalent diode, which would give the same space-charge-free off cathode field intensity as the triode with the grid removed. The plate current obtained with the equivalent diode is the same as that obtained with the triode.

By a suitable conformal transformation of Fig. 1, the triode amplification factor, $\mu$, can be

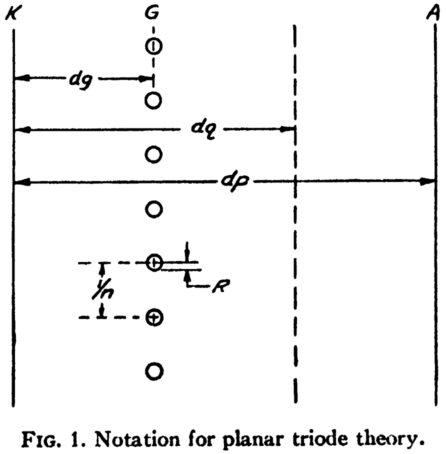


FIG. 1. Notation for planar triode theory.

---

[1] "Microphonic Electron Tube," issued to J. Rothstein, November 27, 1945, U. S. Patent No. 2,389,935.

[2] See W. G. Dow, *Fundamentals of Engineering Electronics* (John Wiley & Sons, Inc., New York, N. Y., 1937), Chapter 2.

determined by electrostatic theory utilizing the quantities enumerated above.[3] The relation is

$$\mu=\frac{2\pi n(dp-dg)-\log_e \cosh 2\pi nR}{\log_e \coth 2\pi nR}$$

and applies with sufficient accuracy when $R<1/10n$.

For $2\pi nR<1$, $\cosh 2\pi nR\rightarrow 1$ and $\log_e \cosh 2\pi nR \rightarrow 0$. Similarly,

$$\coth 2\pi nR=\frac{\cosh 2\pi nR}{\sinh 2\pi nR}\rightarrow\frac{1}{2\pi nR}$$

since $\sinh 2\pi nR\rightarrow 2\pi nR$. Hence, we can write

$$\log_e \coth 2\pi nR=\log_e\frac{1}{2\pi nR},$$

and the above relation becomes

$$\mu=\frac{2\pi n(dp-dg)}{\log_e\dfrac{1}{2\pi nR}}. \qquad (1)$$

For the space-charge-free case with an equivalent diode plate voltage of $Eg+Ep/\mu$ the equivalent diode spacing has been found to be[4]

$$dq=dg\left[1+\frac{1}{\mu}\frac{dp}{dg}-\frac{\mu+2}{2\pi n\mu dg}\log_e \cosh 2\pi nR\right].$$

Under the conditions given above, the last term may be set equal to zero leading to

$$dq=dg\left[1+\frac{1}{\mu}\frac{dp}{dg}\right]. \qquad (2)$$

As an approximation, Child's Law for the space charge limited equivalent diode may be written:[5]

$$Ip=\frac{2.33\times 10^{-3}A}{dq^2}\left(Eg+\frac{Ep}{\mu}\right)^{\frac{3}{2}}. \qquad (3)$$

Where $A$ is the equivalent diode plate area, equal to the actual triode plate area, $Eg$, the triode grid-cathode voltage, $Ep$, the triode plate-cathode voltage, and $Ip$ is the total plate current in milliamperes.

As abbreviations write

$$K=\frac{1}{2\pi n}\log_e\frac{1}{2\pi nR},$$

$$K'=2.33\times 10^{-3}A. \qquad (4)$$

Substituting (1) and (2) into (3) in view of (4) yields

$$Ip=\frac{K'\left[Eg+\dfrac{KEp}{(dp-dg)}\right]^{\frac{3}{2}}}{\left[dg+\dfrac{Kdp}{(dp-dg)}\right]^2}\text{ ma} \qquad (5)$$

for the total plate current of the triode in terms of triode parameters.

The triode transconductance is defined as

$$Sp\equiv\left(\frac{\partial Ip}{\partial Eg}\right)_{Ep=\text{constant}}\text{ in }\frac{\text{ma}}{\text{volt}},$$

where, for the present case, other tube parameters are also held constant.

Carrying out of the indicated differentiation on (5) yields

$$Sp=\frac{\frac{3}{2}K'\left[Eg+\dfrac{KEp}{(dp-dg)}\right]^{\frac{1}{2}}}{\left[dg+\dfrac{Kdp}{(dp-dg)}\right]^2}\frac{\text{ma}}{\text{volt}}. \qquad (6)$$

We wish to compare the effect of a grid displacement for constant $Eg$, $Ep$, and other parameters with (6). Consequently, we define a mechanical transconductance as

$$Mp\equiv\left(\frac{\partial Ip}{\partial dg}\right)_{\text{other parameters constant}}$$

in $\dfrac{\text{ma}}{\text{unit grid displacement}}$.

Carrying out of the indicated differentiation on (5) yields, after suitable rearrangement of

---

[3] F. B. Vodges and F. R. Elder, "Formulas for the amplification constant for three element tubes," Phys. Rev. 24, 683 (1924).

[4] See reference 2, page 44, Eq. (111).

[5] C. D. Child, "Motion of positive ions," Phys. Rev., 32, 498 (1911).

TABLE I. Dimensions of an RCA 6A3.

Radius of grid wire $= R = 0.00225$ in.
Grid wire center spacing $= 1/n = 0.055$ in.
Grid wires per inch $= n = 18.2$.
Grid cathode distance $= dg = 0.045$ in.
Plate cathode distance $= dp = 0.090$ in.
Plate area (4, ½ sections) $= A = 4.22$ in.$^2$.

TABLE II. Tube factors of 6A3. $Ep = +200$ volts, $Eg = -30$ volts.

| Factor | Handbook data | Theoretical | Equation No. |
|---|---|---|---|
| $\mu$ | 4.25 | 3.8 | (1) |
| $Sp$ | $6\frac{\text{ma}}{\text{volt}}$ | $7\frac{\text{ma}}{\text{volt}}$ | (6) |
| $Mp$ | | $3.6\frac{\text{ma}}{\text{in.}\times 10^{-3}}$ grid displacement | (7) |

terms,

$$Mp=Sp\left\{\frac{KEp}{(dp-dg)^2}-\frac{4}{3}\,\frac{\left[Eg+\dfrac{KEp}{(dp-dg)}\right]\left[1+\dfrac{Kdp}{(dp-dg)^2}\right]}{\left[dg+\dfrac{Kdp}{(dp-dg)}\right]}\right\}\tag{7}$$

in milliamperes/unit grid displacement.

From a comparison of $Sp$ with $Mp$, it should be possible to determine the relative effectiveness of the two methods of obtaining plate current variations.

### MAGNITUDE OF THE MOTIONAL TRANSCONDUCTANCE FOR A TYPICAL TUBE

In order to compare the above theory with handbook data, and to determine the magnitude of $Mp$, an RCA 6A3 was selected as a typical tube whose characteristics could be easily determined. One of these tubes was taken apart, and the dimensions appearing in Table I were determined.

The 6A3 is comprised of two tube structures in parallel enclosed in a common envelope. Each structure has an $M$ type filamentary cathode. The grid and plate electrodes are nearly rectangular in form. Consequently, the tube is treated as though it were four planar triode structures in parallel in the above table.

The tube parameters of interest, shown in Table II, were obtained from handbook curves[6] for $Ep = +200$ volts and $Eg = -30$ volt. The voltage values indicated were chosen so that the point on the tube surface under consideration lies on a plane region of the surface. This point also constitutes a compromise between the space-charge-free assumption involved in Eq. (2), and the space-charge-limited case involved in Eq. (3).

The parameters given in Table I were substituted in the relevant theoretical relations, under the voltage conditions mentioned above, and the theoretical values of Table II determined.

In view of the accuracy of the measurements in Table I, and the approximations employed in the theory, the results indicated in Table II with respect to the first two factors are considered satisfactory. We may conclude, therefore, that the value of $Mp$ obtained is correct to within about the same accuracy without experimental verification.

### CONDITIONS FOR THE MOTIONAL TRANSCONDUCTANCE BEING EQUAL TO ZERO

Inspection of Eq. (7) indicates there are several interesting facts concerning the magnitude and sign of $Mp$ under certain conditions. For $Eg$ negative and $Ep$ positive, as is the case for normal amplifiers, the magnitude and sign of the bracket factor vary somewhat as follows. For

$$|Eg|>\frac{KEp}{(dp-dg)},$$

the right-hand term of the bracket factor is negative and large so that the bracket factor is positive and large. Since $Sp$ is positive, $Mp$ is positive and large. For

$$|Eg|<\frac{KEp}{(dp-dg)},\tag{8}$$

the right-hand term of the bracket factor be-

---

[6] RCA Tube Handbook HB-3, Vol. 1-2, Tube Division, Radio Corporation of America, Harrison, New Jersey.

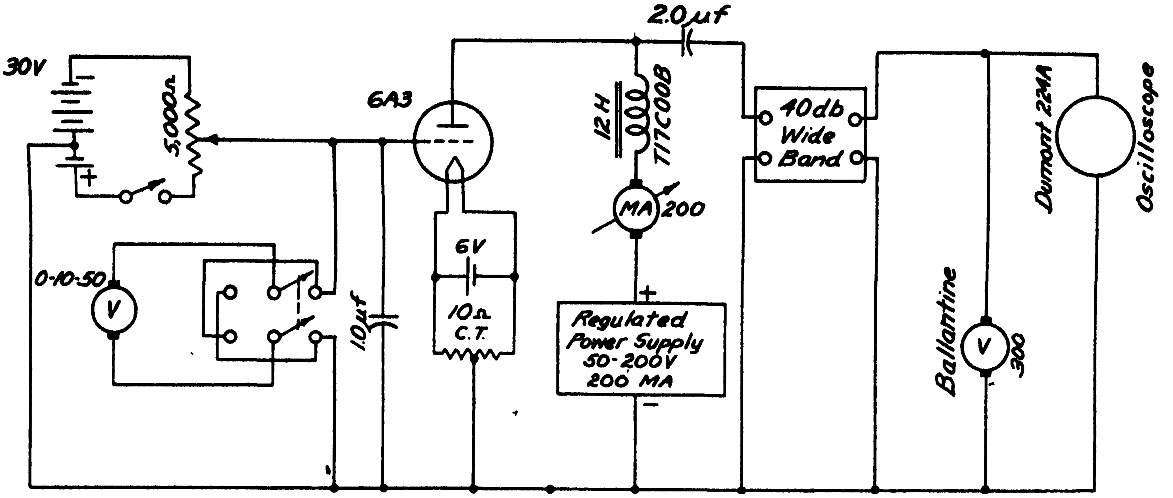


FIG. 2. Schematic diagram of the experimental arrangement.

comes positive and larger as the $|Eg|$ decreases to zero; and as $Eg$ subsequently becomes positive. Hence, the bracket factor passes through zero, when the second term equals the first term, and then goes negative and larger in magnitude as the $|Eg|$ decreases to zero; and as $Eg$ subsequently goes positive. Since $Sp$ remains positive, $Mp$ passes through zero and becomes negative and large in magnitude.

It is well known that, for triodes, $Sp$ increases monotonically as $Ip$ increases. Further, that $Ip$, for fixed positive $Ep$, increases as the $|Eg|$ decreases for $Eg$ negative and continues to increase as $Eg$ goes through zero and becomes positive for normal maximum positive values.

Hence, we can conclude that, in practice, $Mp$ may be zero or small in the normal operating voltage range of triodes where $Sp$ is large so that the tube will still amplify normally as far as electrical fluctuations in grid voltage are concerned.

Since microphonic effects are known to be caused, in great part, by mechanical motion of the grid rods, it is evident that microphonics may be reduced or eliminated by operating a tube, or designing a special tube, under the conditions for $Mp$ to be zero or small.

## EXPERIMENTAL ARRANGEMENT

A simple experiment was set up to check Eq. (7) by mechanically driving a tube and measuring the resulting "microphonic" output signal. Inspection of Eq. (7) indicates that this may be done most easily by varying $Eg$, with $Ep$ fixed, so that the only other variable is $Sp$. This varies in a known manner in terms of $Ip$, which in turn is a known function of $Eg$ for $Ep$ constant.

The schematic diagram of the experimental arrangement is shown in Fig. 2. A 6A3 was operated with battery filament supply to permit

TABLE III. Dominate driving frequencies of microphonic output for a 6A3 and other data. $Ep=+100$ volts; $Eg=-10$ volts; $E_f=6$ volts; volts across speaker $=1.0$.

| Case | Driving frequency for maximum output c.p.s. | Frequencies at which the output is 3 db down from the maximum output c.p.s. | Remarks concerning the output wave form |
|---|---|---|---|
| Transverse Excitation | | | |
| 1 | 194 | 192–202 | Very complex wave form |
| 2 | 310 | 309–317 | Symmetrical complex wave form |
| 3 | 359 | 357–361 | Fair sine wave |
| 4 | 465 | 461–470 | Good sine wave |
| 5 | 549 | 548–550 | Excellent sine wave |
| 6 | 938 | 883–985 | Extremely complex wave form with multiple resonances |
| 7 | 2100+ | 2095–2115 | Good sine wave |
| Longitudinal Excitation | | | |
| 8 | 355 | 351–359 | Complex wave form |
| 9 | 469 | 461–471 | Good sine wave |
| 10 | 544 | 542–546 | Good sine wave |
| 11 | 722 | 719–742 | Extremely complex wave form |
| 12 | 948 | 943–952 | Fair sine wave |
| 13 | 2090 | 2065–2095 | Secondary maximum at 2070 c.p.s. |

hum-free operation. The grid bias was varied by means of the potentiometer across the grid bias battery and its magnitude determined by the grid voltmeter. A.c. pick-up in the grid circuit was avoided by grounding the grid at a.c. by means of the 1-mfd condenser.

The plate supply was a heavy duty, electronically regulated power supply whose output voltage was adjustable over the limits indicated. A voltmeter, not shown in the figure, was connected from plate to ground to indicate the plate voltage. A 12-henry heavy duty choke was used as a plate load in order to obtain a low value of d.c. drop in the plate load over the large range of d.c. plate currents employed.

The resulting "microphonic" output voltage developed across the plate load was impressed, through the 2-mfd d.c. blocking condenser, across the input of a stable, wide band amplifier having a voltage gain of 40 db. The output of the amplifier was indicated by a Ballantine model 300 voltmeter and was also observed on an oscilloscope screen. The no signal output of the circuit was well below the minimum signal measured.

The mechanical arrangement employed is shown in Fig. 3. The output of a sine wave audio oscillator was impressed across the voice coil of a small dynamic speaker. The oscillator output was measured by a Ballantine voltmeter and was held constant throughout the experiment at 1 volt.

Two types of mechanical excitation were employed. As shown in Fig. 3A, the transverse case, the tube was placed on a heavy table top so that the grid planes were parallel with the table top. The speaker was placed on top of the tube envelope so that the speaker diaphragm was pressed tightly against the tube. The speaker was then taped to the table top so that it would maintain this position. It is evident that mechanical excitation of the tube for this case will result in grid rod displacement perpendicular to the plane of the grid.

Figure 3B illustrates the mounting used for the longitudinal case. For this arrangement mechanical excitation of the tube will result in grid rod displacement in the plane of the grid or parallel to the longitudinal axis of the tube.

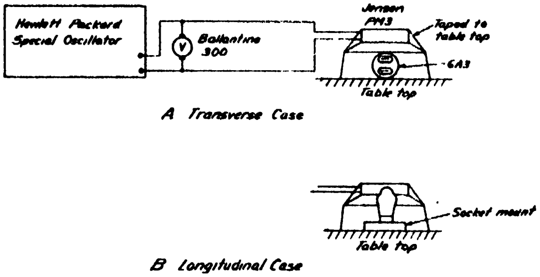


FIG. 3. Mechanical layout of the experimental arrangement.

## EXPERIMENTAL PROCEDURE AND RESULTS

The experimental procedure involved setting the driving oscillator at some constant output voltage and varying the frequency of the oscillator while observing the "microphonic" output on the oscilloscope. The magnitude of the output was highly frequency-sensitive in that the output would vary by as much as 40 db in passing through a frequency range of a few cycles. It is considered that these variations in output are associated with mechanical resonances of the tube structure; and, in particular, with mechanical resonances of the grid rods.

In some frequency ranges the output wave form was greatly dependent on the excitation frequency, indicating complex modes of vibration. In other cases the output wave form was sinusoidal and only the amplitude of the output signal was frequency dependent.

For the particular tube tested, it was found that there were seven dominant frequency ranges for the transverse case and six for the longitudinal case in the excitation frequency range of 100 c.p.s. to 10,000 c.p.s. These are tabulated in Table III below with other relevant information such as a statement concerning the observed wave form and an estimate of the width of the dominate frequency range in terms of the frequencies at which the output signal amplitude, as measured by the Ballantine voltmeter, are 3 db down from the output at the frequency of maximum output. In obtaining these data, the following were held constant: $E_f = 6$ volts, $Ep = +100$ volts, volts across speaker $= 1.0$, and $Eg = -10$ volt. The magnitudes of $Ep$ and $Eg$ were found to effect the output amplitude and wave form, as discussed below, but not the frequency at which the maximum output was obtained.

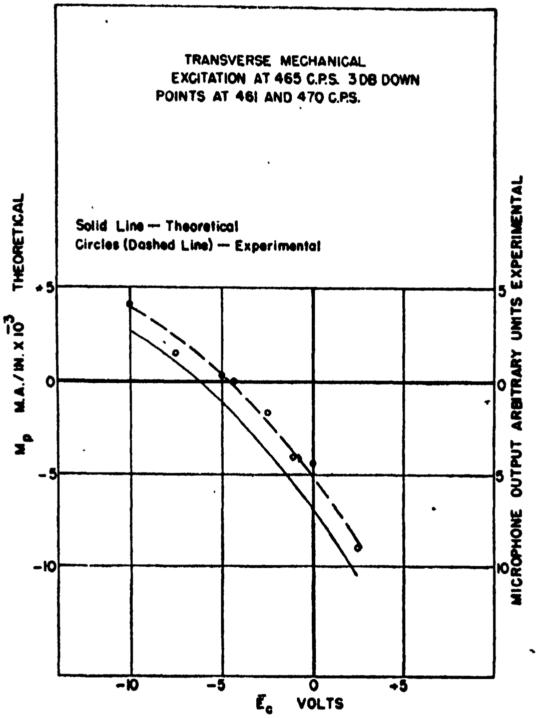


FIG. 4. An experimental check of Eq. (7) with $\bar{E}_c$ as the variable and sine wave mechanical excitation normal to the grid plane.

An experimental check of Eqs. (7) and (8) was then attempted in the following manner. With the driving oscillator accurately set at a frequency of maximum "microphonic" output, as indicated by the output Ballantine voltmeter, the grid bias was varied and the "microphonic" output as a function of grid bias determined. *Ep* was carefully held constant at +100 volts. This value of plate voltage was chosen so that the plate dissipation rating of the tube, 15 watts, was not exceeded at the largest positive grid bias voltage used during the experiment. The latter was +2.5 volts.

Many very interesting factors were noted; some of which are summarized below:

1. With *Eg* varied in the voltage range from −10 volts to +2.5 volts a minimum was noted in the "microphonic" output for the following cases of Table III. For transverse excitation 4 cases, number 1, 3, 4, and 7. For longitudinal excitation 1 case, number 11. This factor is discussed in greater detail below.
2. For the transverse case 2 and longitudinal case 8 the "microphonic" output increased as *Eg* was varied from −10 volts to +2.5 volts in about the same amount that *Sp* increased.
3. For the remaining cases the "microphonic" output decreased in varying amounts as *Eg* was varied from −10 volts to +2.5 volts.
4. For many cases, and in particular for case 6, oscilloscope examination of the output signal indicated that one component of a complex vibration spectra minimized as in 1, while other components did not.

In view of the above, it was assumed that cases 4 and 9 probably represented the gravest mode of vibration of the grid rods so that these cases are studied in detail below. Upon completion of the experiment, an attempt was made to observe the vibration of the grid rods under a low power microscope while the tube structure was vibrated under conditions similar to those of Fig. 3A. This experiment failed since the amplitude of vibration was too small to be detected with the apparatus used.

Considering case 4 of Table III, the experimental results obtained are tabulated in Table IV below.

The results of Table IV are plotted as circles on Fig. 4. The solid line of this figure was obtained by substituting the tube parameters of Table I and the experimental parameters of Table IV in Eq. (7) and solving for *Mp*. To do this, *Sp* was calculated from Eq. (6) under the same conditions.

It is evident that the shape of the experimental curve closely approximates that of the theoretical curve. Considering the type of measurements involved, this is considered a rather good check with the theory in spite of the displacement of the two curves. In addition, the following should be noted:

1. The theoretical value of $Eg(=\bar{E}_c)$ for $Mp=0$

TABLE IV. Experimental results for transverse case with driving frequency of 465 c.p.s. $Ep=+100$ volts; $Eg=\bar{E}c$ = variable; $E_f=6$ volts; volts across speaker = 1.0. 3 db down points 461 c.p.s. and 470 c.p.s.

| $\bar{E}c$ volts | Output volts | Normalized output volts | 11 × normalized output for scale factor | Driving frequency for maximum output c.p.s. |
|---|---|---|---|---|
| −10 | 0.49 | 0.37 | 4.09 | 465 |
| − 7.5 | 0.36 | 0.13 | 1.44 | 465 |
| − 5 | 0.14 | 0.02 | 0.22 | 465 |
| − 4.3 | 0.12 | 0 | 0 | 465 |
| − 2.5 | 0.27 | 0.15 | 1.65 | 465 |
| − 1 | 0.49 | 0.37 | 4.08 | 465 |
| 0 | 0.52 | 0.40 | 4.40 | 465 |
| + 2.5 | 0.94 | 0.82 | 9.00 | 465 |

is −6.2 volts under the conditions of the experiments. This compares with the experimental value of −4.3 volts.

2. When *Mp* changes sign the phase of the "microphonic" output voltage should shift 180°. This was noted experimentally although not indicated on the figure.

3. Most of the residual output signal near and at the minimum value of output is at the second harmonic of the driving frequency.

4. The maximum range in output signal variation greatly exceeded that of the example in some cases. As an example, in case 7 the range of maximum to minimum signal, as a function of $\bar{E}_c$, was over 30 db.

The theory does not cover the case of displacement of the grid rods in the grid plane. Experimentally, this case is probably complicated by phase relations between the vibrations of adjacent grid rods and by transverse components of the resulting vibrations. Experimental results for this case, case 9, are plotted in Fig. 5 as circles to illustrate the results obtained. As noted, the driving frequency is nearly the same as for case 4. The absolute magnitude of the output signal for $Eg = -10$ volts was nearly the same in both cases. This is not considered significant, however, since there is not reason to believe that the driving forces for the two cases are the same.

## CONCLUSIONS

1. The method of utilizing the displacement of the grid relative to the rest of the tube structure to obtain plate current variations which are a function of the displacement, and hence of the mechanical force causing the displacement, is feasible from an electronic standpoint.

2. Microphonism of tubes, under certain conditions, may be reduced as indicated by the theory developed and in terms of two factors or a combination thereof:

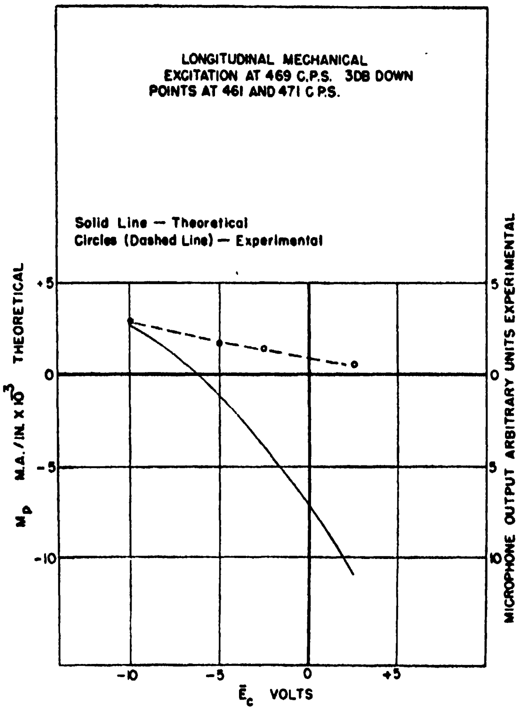


FIG. 5. Similar to Fig. 4 with mechanical excitation approximately in the grid plane.

(a) By designing a tube for minimum *Mp* given the operating potentials or (b) for a given tube to select the optimum operating potentials so as to obtain a minimum *Mp*.

In view of the latter, it may well be advantageous to operate the input stage of a high gain automatic volume controlled amplifier with fixed bias to reduce microphonics even though the flatness factor of the system may be decreased.

The above is thought to be interesting and useful, but from the practical standpoint further theoretical work should be done for the cylindrical triode and multi-grid tube cases. The construction of a planar triode whose grid can be moved relative to the cathode and plate should also be attempted. It is hoped that further work along these lines can be carried out in the future.

# The Effect of Hydrostatic Pressure on the Fracture of Brittle Substances

P. W. BRIDGMAN
*The Physics Laboratory, Harvard University, Cambridge, Massachusetts*

(Received November 11, 1946)

An arrangement is described permitting the fracture of brittle materials under the action of tensile stress superposed on hydrostatic pressure. The hydrostatic pressures range up to 30,000 kg/cm$^2$. The tensile stress superposed on the pressure required to break Pyrex glass is a strong function of the material by which pressure is transmitted to the lateral surface of the glass, but in all cases of fracture produced at pressures above 25,000 kg/cm$^2$ the net stress at fracture was compressive, that is, tensile fracture takes place against the direction of the stress. Beryllium and phosphor bronze lose the brittleness which characterize them at atmospheric pressure and fracture in tension under pressure after marked plastic deformation. Carboloy remains brittle in tension under pressure, but its tensile strength may increase by a factor of three. NaCl elongates plastically in tension under pressure with no obvious disturbance of the optical homogeneity. Pipestone remains completely brittle under pressure and of very low strength. In simple compression combined with hydrostatic pressure Pyrex glass increases greatly in strength. Single crystal $Al_2O_3$ under the same conditions exhibits plastic slip on the basal plane without fracture. In the discussion the importance of the condition of energy release as a factor controlling fracture is emphasized. The Griffith conception of fracture as resulting from stress concentration at the ends of microscopic cracks is consistent with many qualitative aspects of the situation, but probably cannot be carried through quantitatively.

## INTRODUCTION

IN a number of previous papers[1] I have described the effect of hydrostatic pressure in increasing the ductility of substances like mild steel under the action of tensile forces. In this paper the experiments are extended to a number of materials ordinarily brittle, like glass.

The effect of pressure is investigated on both compressive and tensile strength. Measurements of compressive strength may be made straightforwardly, but the tensile tests require a special technique. Under ordinary conditions there are technical difficulties in making tests on substances like glass with the conventionally shaped tension specimen, because of the effect of stress concentration at the corners of the specimen, resulting in capricious fracture. Most fortunately it turns out that a technique which is simply adapted to the application of tension under hydrostatic pressure at the same time uses specimens of such a shape that the difficulty from stress concentration disappears, making the investigation of the tensile properties under pressure much easier than at atmospheric pressure.

The experiments on tension to be described in the following are in a certain sense an extension of experiments made long ago on what I called the "pinching-off" effect.[2] In these experiments rods of glass or other brittle substances passed completely through the pressure vessel, emerging through stuffing boxes, the part within the vessel being exposed to hydrostatic pressure. On raising pressure high enough, the glass rod fractured by a clean tensile break at some place inside the pressure vessel. These experiments were mostly qualitative in nature, and no attempt was made to get consistent values for the pressure required to produce fracture. The chief result from these experiments was to emphasize that it is possible to have tensile fracture across a plane across which there is no stress component. This is, from the point of view of some of the traditional conceptions of fracture, a highly paradoxical thing. The conclusion which I drew from these experiments has, therefore, understandably been met with considerable scepticism in some quarters, and with the attempt to explain away the phenomenon as caused by various secondary effects. For a long time I have had it in mind to make a more exact study of the "pinching-off" effect. A few years ago I found, in some unpublished experiments, that the precise pressure at which the "pinching-off" fracture is produced in glass is

---

[1] P. W. Bridgman, J. App. Phys. 17, 201 and 692 (1946). Rev. Mod. Phys. 17, 3 (1945).

[2] P. W. Bridgman, Phil. Mag. p. 63 (July, 1912).

extraordinarily sensitive to surface conditions. This indicated the necessity for a more elaborate investigation of the whole effect. With these new experiments, this extended examination is now possible. We have now two degrees of freedom in the stress system, whereas formerly there was only one. Formerly the set-up was such that the longitudinal component of stress was continually zero, and the two equal lateral components were increased until fracture occurred. Now, the longitudinal component and the two equal lateral components may be independently varied, so that fracture may be studied over a much wider range of conditions. The paradox that characterized the "pinching-off" effect is now carried much further, for now tensile fractures can be produced across planes on which the stress is a *compressive stress* so high as to be beyond any possible experimental error. Thus. to mention only a single example, tensile fracture has been produced in glass under a mean hydrostatic pressure of 26,000 kg/cm$^2$ with a superposed tension of 11,000 kg/cm$^2$, so that the net stress across the plane on which the tensile break occurred was a compressive stress of 15,000 kg/cm$^2$.

## THE METHOD

The method and the general nature of the problems encountered may be indicated by means of a highly idealized mental experiment. Figure 1 represents a compound tension specimen consisting of a central cylinder *A* of the brittle substance to be tested, and of two "pull" pieces *B* with shoulders and the same small diameter as the specimen, mounted together in the thin cylindrical sheath *C* which is attached by solder or otherwise to the pieces *B* at *D* in such a way as to be liquid tight. The whole compound specimen is immersed completely in a liquid, to which any desired hydrostatic pressure is applied, and, superposed on the hydrostatic pressure, a tensile force is then applied to the pull pieces *B*. The situation is idealized by supposing that the pull pieces are infinitely strong, that the lateral contraction of *A* and *B* under the complete stress system is the same, so that their diameters continue to be the same, and that the sheath *C* has negligible longitudinal strength, offers zero friction on its inside surface, and is infinitely stiff so that it cannot be pushed by the external pressure into any cracks which might be formed on the surface of *A* or *B*. We now inquire what will happen if the pressure is first raised to some high value, at which it is kept constant, and the tension is then increased indefinitely. It is evident in the first place that the compound specimen cannot separate at the surface of separation between *A* and *B* as long as the tension is less numerically than the pressure. For if incipient separation should start at this surface, the surfaces of separation would be pushed back into contact by the pressure, which is greater than the tension. All the more, if separation cannot occur at surfaces already free between *A* and *B*, it cannot occur in the material of the solid rod *A*. That is, under a hydrostatic pressure, *P*, the tensile strength of the material *A* is, under these conditions, at least as high as the pressure *P*. Ideally, it seems impossible to place any limit to this behavior, so that by increasing indefinitely the hydrostatic pressure the tensile strength may also be indefinitely increased. This means that the numerical value of the pressure coefficient of tensile strength approaches unity under these conditions. This in itself is highly anomalous. It

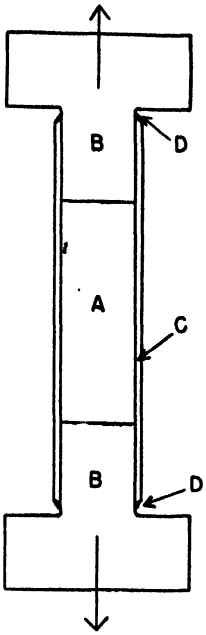


FIG. 1. Tension specimen, *A*, with steel pull pieces, *B*, and sheath, *C*, as used in combining hydrostatic pressure and tension.

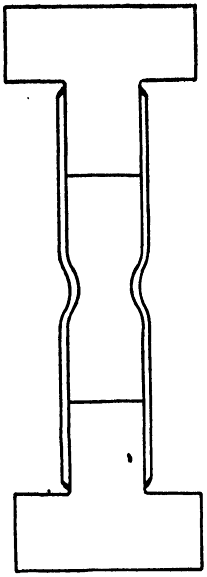


FIG. 2. Modification of arrangement of Fig. 1 in which the tension is concentrated in a neck, permitting the fracture in tension of the stronger materials.

is a matter of experience that the pressure coefficient of most physical properties of substances like glass is of the order of magnitude of perhaps $10^{-5}$, if the unit of pressure is taken as the kilogram per square centimeter. Yet here we have the necessity for a coefficient $10^5$ times as high. The situation is also anomalous when considered from the point of view of strain. Under combined tension and pressure the specimen $A$ elongates, and we have apparently the necessity for an indefinitely great elongation with no fracture. To say that beyond a certain point the specimen will receive plastic set on release of stress does not relieve the paradox of the situation while the stress is in action.

This idealized experiment, leading to such paradoxical conclusions, may be made the basis of the actual experiment, and also suggests the direction in which the experimental results may be expected to go, which is indeed the direction of paradox. In practise the pull pieces $B$ may be made of heat-treated steel, and the specimen $A$ made of the brittle substance under examination, such as glass or brittle alloy or what not. The sheath $C$ may be made of copper, a few thousandths of an inch thick. We now see at once that we have a method of applying tension to a brittle specimen without shoulders or regions of stress concentration on the specimen itself. The pull pieces $B$ may be re-used from experiment to experiment, so that the preparation of the specimen could not well be simpler. In the simple form shown in Fig. 1 any tension can be applied up to the magnitude of the pressure itself, assuming that there are no limiting complications in practise. These complications do, in fact, occur, since the copper is not infinitely rigid, and would eventually blow into the surface of separation between $A$ and $B$ before tension reaches its theoretical upper limit. This difficulty may be avoided by artificially fashioning a neck on the tension specimen, as shown in Fig. 2. The neck may be easily made of such proportions that there is no harmful stress concentration in the neighborhood. By making the neck of suitable dimensions, any tensile stress may be applied in this manner, without grips, to any material up to the fracture point, whether that is above or below the surrounding hydrostatic pressure. This arrangement has been used, for example, in testing the tensile strength of Carboloy under pressure, the tensile strength being much higher than the highest hydrostatic pressure that was applied.

The material of the sheath is a vital factor in the result, for not only may it be blown between the surface of separation of $A$ and $B$, but it may also be blown into any incipient cracks which appear in the surface of the specimen $A$. One would anticipate, therefore, that the measured tensile strength would be a strong function of the surface conditions, as exemplified by the material of the sheath. This did indeed prove to be the case. Various materials were tried. In practise the sheath sometimes was made compound, still keeping copper over the surface of separation of $A$ and $B$, but replacing the central part by some other substance such as lead or rubber, which must naturally be attached in such a way as to prevent access of liquid. It would have been desirable if conditions could have been pushed to the limit by permitting contact between the liquid and the specimen all over its lateral surface, while preventing entry of the liquid into the interface between $A$ and $B$, but no feasible way of doing this presented itself. As it was, a sufficiently wide range of external conditions was attained by the use of copper, lead, and rubber to indicate the general nature of what might be expected.

The arrangement indicated in Figs. 1 and 2 has points of resemblance to an arrangement probably first used by Richart, Brandtzaeg, and Brown[3] in studies of the effect of compound stresses on concrete and later by Balsley[4] working with Griggs on minerals. Both these investigators used straight cylindrical steel pull pieces, attached to the specimen by a sheath. The pull pieces were taken out of the pressure chamber through stuffing boxes, and on the outside were subjected to an independent compressive stress, which, as it varied from greater to less than the internal hydrostatic pressure was capable of exposing the specimen to a superposed stress either of compression or tension. There are technical difficulties with this arrangement, however. The stuffing boxes have to be accurately aligned and are not capable of as wide a pressure range as that used here. The maximum pressure in the Illinois experiments was 300 kg/cm², and in those of Balsley 10,000 kg/cm². Neither of these experimenters used necked specimens in order to permit raising the tension to values exceeding the hydrostatic pressure, but there is no reason why necked specimens should not be used with externally projecting pull pieces.

During the war Gurney and Rowe[5] in England made a large number of measurements of the pinching-off effect (which they describe as fracture by "radial" pressure) on glass and several plastics, and in a number of instances used necked specimens, thus permitting any desired tensile stress depending on the dimensions of the neck. These experiments were limited, however, because of the extension of the specimen itself through stuffing boxes, to hydrostatic pressures not greater than the "pinching-off" pressure of the straight cylindrical specimen.

In addition to the effect of one-sided tension, the effects of one-sided compression were also studied under pressure. These experiments were perfectly straightforward, and the details need not be specially indicated, for there is no such complication as with the tension specimens, but the stress may be applied to simple cylindrical specimens by compressing between platens. The specimen may be imbedded in lead or other suitable soft material when it is desired to prevent access of the transmitting liquid.

The various other details of the experiment, such as methods of applying and measuring the force, were the same as in my previous experiments, and the details have already been sufficiently described. An improved method of measuring the deformation was used in these present experiments, which has not yet been described in print. This employed a new insulating plug with seven terminals and a potentiometer method, which permitted measurement of the changes of dimensions from the inside of the pressure vessel, instead of from the outside, as hitherto. However, since these deformations during the action of stress are to play no role in the following, the description of this new arrangement can be deferred until the compressibility measurements in which it played an essential role are described.

## THE EXPERIMENTS

### Tensile Tests

#### *Glass*

The experiments on glass were made on Pyrex, all specimens being taken from the same batch. Five experiments were made in all. In the first a copper sheath was used 0.016 inch thick over the junction between steel and glass, turned down to a thickness of 0.005 inch over the central part of the glass. No extravagant precautions were taken with the glass, which was 0.44 inch long and 0.12 inch diameter. The lateral surface was the natural surface of the rod as provided commercially. The ends were lapped by hand with a fine carborundum stone perpendicular to the axis. Care was taken not to chip the edges. This specimen was exposed to a pressure of 27,000 kg/cm², and then a superposed tension added up to 19,000 kg/cm², when separation occurred between glass and one steel pull piece. The glass was undamaged, as was found by dissolving off the copper sheath in acid. The pull pieces were found to be too soft, and under this tension received a permanent elonga-

---

[3] F. E. Richart, A. Brandtzaeg, and R. L. Brown, Bull. No. 185 (1928), Eng. Exp. Stat. Univ. Ill.

[4] J. R. Balsley, Trans. Am. Geophys. Un. Part II, 519 (1941).

[5] C. Gurney and P. W. Rowe, Royal Aircraft Establishment, Farnsborough, Rep. No. Mat. 5 (May, 1945), and Rep. No. Mat. 8 (November, 1945).

tion of 1.5 percent. This flow of the steel may have been responsible for the separation. The glass received no measurable permanent change of dimensions.

In the next experiment, copper ferrules were used over the surface of separation of glass and steel, but the central part of the glass was covered with a soft Neoprene tube, extending over the copper ferrules. At a hydrostatic pressure of 25,000 kg/cm$^2$ and a superposed tension of 5900 kg/cm$^2$, making the net compressive stress in the glass 19,100 kg/cm$^2$, this failed by a tensile break at the edge of the copper ferrule. On removing the ferrules from the glass, the glass fell apart into several disks separated from each other by clean tensile breaks. In this experiment the edges of the ferrules were comparatively thick, and it seemed probable that the fracture was in some way connected with a discontinuity of stress at the edge. The experiment was accordingly repeated, but with the edges of the copper ferrules feathered down to vanishing thickness. At a hydrostatic pressure of 25,300 kg/cm$^2$ and a superposed tension of 11,400, failure occurred at the interface. The entire central portion of the glass, where it had been exposed to the Neoprene, was separated into thin disks, still hanging lightly together. In the fourth experiment, copper ferrules again guarded the surface of separation of glass and pull pieces, and the central part of the glass was covered with a heavy lead sheath, soldered to the copper, The central part of the glass was necked down by grinding and polishing from .120 to 0.106 inch in order to concentrate the tensile stress at the center and locate the fracture. Under a hydrostatic pressure of 28,000 kg/cm$^2$ this broke in clean tensile fracture across the smallest diameter of the neck under an average superposed tensile stress at the neck of 15,500 kg/cm$^2$. Tensile fracture, therefore, occurred here against a net compressive stress of 12,500 kg/cm$^2$. The lead was melted off, and the copper dissolved off in aqua regia. An additional tensile break appeared on thus stripping the specimen, and the entire mass of the remaining glass was copiously fissured with fine disking cracks, the glass still hanging lightly together. In the fifth experiment, the glass was necked down at the center from .120 to 0.080 inch by grinding, polishing, and fire polishing to remove the last trace of mechanical cracks. The glass, together with the pull pieces, was mounted in a heavy straight copper sheath 0.21 inch in diameter, which was then collapsed tightly around the neck in the glass by a preliminary exposure to 8000 kg/cm$^2$, and then turned down to a thickness of 0.005 inch over the neck. Under a hydrostatic pressure of 27,000 kg/cm$^2$, this failed by a clean tensile break across the neck at an average superposed tension at the neck, corrected for the strength of the copper, of 24,500 kg/cm$^2$. This makes fracture to have occurred at a net compressive stress of 2500 kg/cm$^2$.

### *Carboloy*

Tests were made on Carboloy of grade 999, which is the grade containing the smallest quantity of binder furnished by the General Electric Company. To increase the stress the specimen was necked by grinding with diamond. Two unsuccessful preliminary experiments were made in which the strength of the Carboloy was not estimated high enough and the neck diameter accordingly was not made small enough. In two final experiments the area at the neck was made somewhat less than one quarter the area at the ends. In the two final experiments hydrostatic pressure was raised to 27,000 kg/cm$^2$. In the first of the two, there was a leak at the maximum load, suddenly throwing an uncontrolled and unmeasured increment of tension on the sample, under which it fractured. The best estimate of the superposed tensile load at the neck at fracture was somewhere between 56,000 and 60,000 kg/cm$^2$. In the second, there was no mishap and tensile fracture occurred at a superposed tension at the neck of 54,500 kg/cm$^2$. In both cases the break was a clean fine-grained tensile break square across the narrowest part of the neck, with no evidence whatever of any plastic flow in the neck, and no trace of the little shearing rim around the edge of the break that is almost always found in hardened steel.

The tensile strength under 27,000 is thus between two and three times as high as any value that has ever been recorded for any grade of Carboloy at atmospheric pressure by the method of the bending of slender rods. It was a surprise that there was no evidence of plasticity. Under one-sided compressive stress this grade of

Carboloy permits plastic shortening up to at least 10 percent.

### *Beryllium*

Beryllium as ordinarily obtained commercially, even in the purest form, breaks at atmospheric pressure in tension with no appreciable elongation and with tensile strength not more than 2000 kg/cm². Being a crystalline material it was anticipated that under hydrostatic pressure it might exhibit a degree of plasticity. The first experiment with beryllium, which was kindly furnished by Professor John Chipman of the Massachusetts Institute of Technology, employed steel pull pieces, by the same technique as used for glass and Carboloy. The specimen was necked to 0.7 of the full diameter, and mounted in a copper sheath, collapsed around it as in the last experiment on Pyrex. Under a pressure of 27,500 kg/cm² this broke at the neck with a further reduction of area at the neck of 48 percent, thus showing the anticipated ductility under the action of pressure. The neck at the break was very jagged, and the contour was finely corrugated around the periphery in an unusual way. The superposed tensile stress at fracture was 7900 kg/cm², on the original area of the neck. Two other experiments were made with beryllium, in which the conventional shaped tensile specimens were used, like Fig. 1, entirely fashioned out of a single bar. Mr. Charles Chase skillfully found how to fashion the beryllium rod with a Carboloy tool. In these two experiments the beryllium was directly exposed to the action of the pressure transmitting liquid, a mixture of iso-pentane and commerical "pentane." In the first experiment the specimen pulled apart at the shoulder under a hydrostatic pressure of 23,000 kg/cm². This was evidently the effect of stress concentration. There was, however, measurable extension in the body of the specimen. In the second specimen, the fillet at the shoulder was given a larger radius and the center of the specimen was further necked down to localize the break. Under a pressure of 18,000 kg/cm², this broke across the center of the neck with a coarse granular break, with a 20.6 percent reduction of area, and a maximum superposed tensile stress of 5400 kg/cm² calculated on the original area.

There would seem to be no reason to think that the transmitting liquid exercises any effect in the case of this material. It would appear that increasing hydrostatic pressure progressively raises the ductility and the tensile strength.

### *Phosphor Bronze*

For commercial welding a bronze is used consisting of 93 percent copper and 7 percent phosphorus. Under atmospheric conditions this material as ordinarily supplied in the cast condition is coarsely crystalline and completely brittle. It was of interest to find whether ductility is imparted by pressure. Three experiments were made. The specimens were made in one piece from the commercial cast rod, and were exposed to direct contact with the pressure liquid. The first two of these were not pulled to fracture, the ductility proving to be greater than anticipated, so that the dimensions originally given did not permit extension to fracture. The first of these was pulled under 22,000 kg/cm² to a reduction of area of 12.6 percent, and the second under 29,000 kg/cm² to 18.5 percent reduction. The elongation of both these specimens was uniform, necking not having yet started. The maximum superposed tensile load on the second specimen was 8700 kg/cm². In the third experiment the original specimen was turned down further over the central portion, to locate the break. Under a hydrostatic pressure of 29,900 kg/cm² this broke with an 80.2 percent reduction of area. The break was a burnished shearing break at approximately 45° across the narrowest part of the neck. The load at fracture was not obtained.

A control run on a similarly shaped specimen at atmospheric pressure yielded, as expected, a completely brittle fracture at the bottom of one fillet.

### $Al_2O_3$ (*Synthetic Sapphire*)

The Linde Air Products Company has recently developed a method of making single crystal rods of $Al_2O_3$, which have found various industrial uses. One of these rods was tested in tension, and I am much indebted to the Linde Company for making the specimen available. The hexagonal axis was oriented, as determined by the Linde laboratory, at an angle of 20° with the length of the specimen. It was mounted with steel pull pieces, copper ferrules, and Neoprene over the

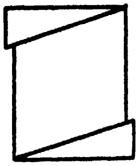


FIG. 3. Schematic indication of the manner of fracture of Pyrex glass in simple compression when supported by hydrostatic pressure.

central part. The central part was necked down by grinding with fine diamond powder from an original diameter of 0.100 to 0.070 inch, in order to locate the break. Although the final grinding was with the finest powder on hand, about No. 500, fine scratches were visible on the surface of the ground part. This material as ordinarily supplied is by no means as perfect as the crystals which are formed in nature, but usually there are imperfections visible to a low power glass in the form of minute bubbles irregularly scattered through the interior. The specimen used for this test was unusually perfect in this respect, however, and showed no such visible imperfections. This was pulled at a pressure of 24,700 kg/cm$^2$. At a superposed tensile load of 4900 kg/cm$^2$ it fractured on a plane perpendicular to the axis, not at the neck, nor at the edge of the copper ferrule, but at a place where the diameter was the full original diameter, situated about half-way between the edge of the ferrule and the neck. The conclusion forced by this is that the material is very inhomogeneous. The neck withstood without fracture a stress of 10,200 kg/cm$^2$ at the same time that the ruptured part was giving way under 4900. The surface of the break itself was not homogeneous. About half the area of the break was a mirror-like plane, with fine ripple marks around the edge, but the remainder of the section had a mat appearance, perhaps caused by the presence of bubbles too small to see with the hand glass.

### *Pipestone* (*Catlinite*)

This material, an iron-rich clay-stone, has a typical composition: $SiO_2$ 57.4, $Al_2O_3$ 25.9, $Fe_2O_3$ 8.7, $H_2O$ 7.4, MgO and CrO trace. Mechanically it seems to be an exceedingly fine-grained cemented aggregate. I have used it extensively for the insulation of electrical leads into the high pressure chamber, and in this usage it is capable of supporting high shearing and compressive loads without appreciable flow. When it does yield in this usage it disintegrates to a fine powder. Advantage was taken of the new technique to study its behavior in tension, it being quite impossible to make the conventional tension tests at atmospheric pressure. The specimen used was a single straight cylinder 0.180 inch in diameter and 0.438 inch long, mounted between steel pull pieces in a copper sheath turned to a thickness of 0.005 over the central part. At a pressure of 27,400 kg/cm$^2$, this failed on a clean tensile plane well within the central portion of the specimen, and with a reduction of diameter at the break of not more than 0.0005 inch. The superposed tension at fracture was only 430 kg/cm$^2$. There was no evidence of the copper sheath being blown into the fracture, but the copper was cleanly sheared through at the break. In view of experience with other materials, such a negligibly small effect of pressure on tensile properties was hardly anticipated. It is doubtless connected with the unusual mechanical composition of this material.

### *NaCl*

Under normal conditions at atmospheric pressure rocksalt breaks brittlely in tension on one of its numerous cleavage planes. It is well known, however, that under special conditions, as when properly supported or when in aqueous solution, it may be made to support plastic deformation. Two tests were made of the effect of hydrostatic pressure. Both specimens were cut from the same natural crystal, with one of the natural cleavage planes perpendicular to the length. They were used in the form of straight cylinders, mounted between steel pull pieces. The first was used with a copper sheath 0.007 inch thick over the central part, and the second with copper ferrules and lead over the central portion. The first was pulled under a hydrostatic pressure of 29,800 kg/cm$^2$ to a reduction of area of 20 percent under a superposed "true" tension as calculated on the final diameter, of 520 kg/cm$^2$. At the necked part there was no loss of optical homogeneity—no evident slip planes or other evidence of crystal structure, and the cross section remained round. The second specimen, with lead sheath, was pulled under 29,300 kg/cm$^2$ pressure. The experiment was terminated by leak, which prevented the measurement of the tension, but the specimen necked down to a reduction of area of 14.6 percent. Again the specimen retained its perfectly homogeneous appearance.

Voigt[6] experimented on the effect of gas pressures up to 60 kg/cm² exerted by $CO_2$ on the tensile strength of NaCl. He found that up to this pressure the superposed tensile load required to fracture was independent of pressure, and that the fracture continued to be a clean brittle break with no trace of plastic flow. Break occurred on the cleavage planes, and the superposed stress to cause fracture was 57 kg/cm². Evidently his pressures were not high enough to bring about the plasticity which is exhibited at higher pressures. It is to be noted that Voigt's tensile strength was only one-tenth of the stress which at high pressure produced only flow without fracture.

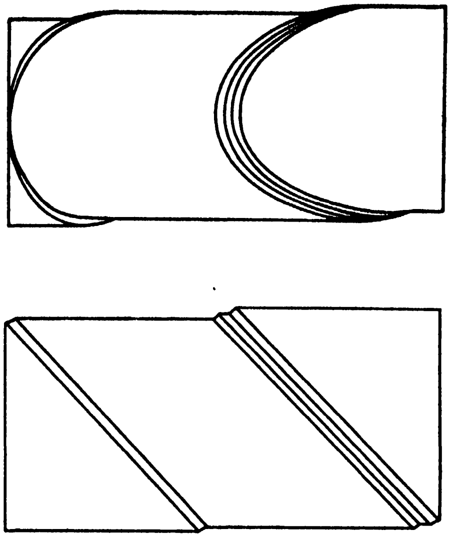


FIG. 4. The slip planes developed in single crystal $Al_2O_3$ under simple compressive stress when supported by hydrostatic pressure.

## Compressive Tests

### *Glass (Pyrex)*

Two experiments were made on glass of the same origin as that of the tension tests. The first specimen was 0.235 inch long and 0.125 inch in diameter. It was compressed in the regular way between Carboloy platens, after soldering into a lead sheath to prevent contact with the pressure liquid. Compressive load was applied at a hydrostatic pressure of 28,000 kg/cm². At an additional compressive load of 18,900 kg/cm² the capacity of the grid to measure compressive stress was reached, and the experiment had to be discontinued. The glass was found undamaged except for a slight chipping around one edge. In the second experiment, the specimen was made smaller, 0.131 inch long and 0.069 inch diameter. The method of mounting was as before. Compressive load was applied under a hydrostatic pressure of 28,000 kg/cm². The specimen failed under a superposed compressive load of 47,000 kg/cm², calculated on the original area. The lead sheath maintained the fragments in approximate position; when the sheath was removed the specimen was found with the appearance suggested in Fig. 3. Two wedge-shaped regions on the ends had slipped sidewise on planes at an angle with the axis of about 65°; the material in these wedges was completely comminuted to a fine powder, opaque and white. The central portion was filled with conchoidal fracture surfaces; the fragments were much larger than at the ends and were transparent.

It is to be noticed that the effect of hydrostatic pressure is markedly greater in increasing the compressive strength of glass than in increasing its tensile strength.

### *$Al_2O_3$*

A number of tests were made on material of the same origin as that of the tensile tests. The results on this material were very capricious, and would hardly merit description here if it were not for one remarkable feature. Four specimens were tested at atmospheric pressure as controls. The crushing strengths of these were 40,000, 8000, 12,000, and 38,300 kg/cm², respectively. Three specimens tested under hydrostatic pressures of 23,800, 25,300, and 22,200 kg/cm² failed completely at additional superposed compressive loads of 18,300, 7700, and 9200 kg/cm², respectively. The failure of these specimens under pressure was remarkable in that it was not catastrophic, but was preceded by a long period of yield under approximately constant compressive load, as in a tension specimen that shows

---

[6] W. Voigt, Göttinger Nachrichten 521 (1893).

high plastic extension. Another experiment was, therefore, made in which the experiment was terminated before the final failure, so as to permit examination of the specimen. This was exposed to 24,000 kg/cm² hydrostatic pressure and a superposed simple compression of 6300 kg/cm². The appearance was as indicated in the sketch, Fig. 4, which I owe to the kindness of Mr. McCandless, of the Linde Air Products Company. The slip planes, which were inclined at approximately 45° to the compression axis, were identified by Mr. McCandless as the basal 0001 plane of the crystal. This is *not* the plane of cleavage when a sapphire rod is broken at a neck by tension-plus-bending. The cleavage plane under these conditions is $10\bar{1}1$. It may be that in this case fracture occurred on 0001 because it by accident happened to coincide with the plane of maximum resolved shear. It would seem, therefore, that under sufficiently high supporting hydrostatic pressure internal slip without fracture is possible in a *crystalline* substance as brittle under normal conditions as sapphire. The crystal structure obviously plays a decisive role; no tendency to plastic flow has ever been found in glass.

The maximum compressive strength found under pressure in the experiments above is much less than that previously found under pressure for $Al_2O_3$ of other origin.[7] For this, compressive strengths as high as 70,000 kg/cm² have been observed. The difference is doubtless connected in some way with fine scale inhomogeneities. There seemed to be no correlation between the compressive strengths obtained in these new experiments and the crystal orientation.

*Note added on reading proof January 17, 1947.*

I have just found similar slip bands in *tension* on the basal plane, at angles of 45° and 30° to the length, when the specimen is enclosed in a copper sheath.

### *Sintered Carbides*

A number of observations have been made on these materials incidentally in the course of explorations for materials best suited for the construction of high pressure vessels. It has already been stated that Carboloy (tungsten carbide cemented with cobalt) will support plastic shortening in simple compression when supported by high hydrostatic pressure. The pistons of the piezometers used in measuring compressions to 100,000 kg/cm² are made of Carboloy, and because of this, considerable work has been done in examining the behavior of the piston material. In general the grades of Carboloy containing small percentages of cobalt are best suited for pistons. I have used the grades known as 905, and lately exclusively 999. Shortenings in simple compression up to 10 percent have been observed at hydrostatic pressures around 30,000 kg/cm². Carboloy does not work harden when it yields plastically in compression, but the shortening continues proportionally to time. Compressive stresses as high as 150,000 kg/cm² have been obtained in Carboloy for a short time. The rate of yield under these conditions is, however, so high as to make the use for pistons infeasible. My pistons in the compression measurements have seldom been carried above 110,000, and more usually are not carried above 105,000. Under these stresses the total plastic shortening of the pistons is something of the order of one quarter of a percent for a duration of maximum stress of fifteen minutes. There is considerable variation in piston material of ostensibly the same composition, and it pays to make a preliminary examination in order to pick out the best pieces.

It is not inconceivable that other materials than Carboloy would be better adapted for standing high compressive stress when supported by hydrostatic pressure, even though the material might not be so well adapted for ordinary conditions of use. The natural direction in which to seek for improvement is in diminishing the amount of binder, since it would appear that the plastic flow takes place in the binder. A cursory examination was made of some of the more obvious or easy possibilities, but with no improvement. In fact, none of the materials tried approached Carboloy in performance; doubtless we went too far in the direction of minimizing the binder. In this exploration I was fortunate to obtain the cooperation of Mr. deWald, who has specialized in the subject, and who prepared the specimens for me at Massachusetts Institute of Technology, where he enjoyed the facilities of the Department of Metallurgy. The following materials were tried with the following results.

---

[7] P. W. Bridgman, J. App. Phys. 12, 461 (1941).

(1) A mixture of 97 percent TaC, 2 percent VaC, and 1 percent $Mo_2C$ with no binder. Under a pressure of 23,000 kg/cm$^2$ this failed with a completely brittle fracture under a superposed compressive stress of 36,000 kg/cm$^2$.

(2) WC with 0.25 percent Co as binder. Under 25,000 kg/cm$^2$ this failed brittlely under 47,500 kg/cm$^2$ superposed compressive stress.

(3) TaC with 0.25 percent Ni as binder. Under 25,000 kg/cm$^2$ hydrostatic pressure this failed brittlely under 30,700 kg/cm$^2$ superposed compressive stress.

There is room for much more work on this subject. In particular, systematic investigation should be made of the effect of varying the grain size and the time and temperature of sintering.

## DISCUSSION

This should be regarded as a first paper on this subject, in which the general nature of the effects is suggested; there is obviously need for much more investigation in detail, both with a wider range of materials and a wider range of stress conditions.

Perhaps the most important single result is to emphasize the reality of a condition of fracture which may often be disregarded. This is: fracture cannot occur unless the process of fracture is an energy releasing process. The reason that in the idealized experiment with which we started our exposition tensile fracture could not occur until the superposed tension reached a value at least equal to the ambient hydrostatic pressure was merely that fracture under these conditions was not an energy-releasing process. For if fracture were to occur, the work done against the hydrostatic pressure would be greater than the work received from the force producing the tension. The fact that fracture occurs in the actual experiment under tensions less than the hydrostatic pressure, that is, against a net compressive stress, is a consequence of the difference between the ideal and the actual case. This difference is in the lateral conditions. In the ideal case, there was an infinitely rigid sheath shielding the lateral surface from entry by the pressure liquid if a crack should appear. In the actual case, the sheath is not infinitely rigid, and if the external pressure is high enough, the sheath may be forced into incipient cracks, and this is an energy-releasing process. According to this view, tensile fracture of a brittle substance in a medium carrying hydrostatic pressure starts on the external surface. This view is confirmed by the extreme sensitiveness of tensile fracture under these conditions to the surface conditions. Thus in the case of glass, we may have fracture by the "pinching-off" effect at a pressure of 1500 kg/cm$^2$ if the external medium is water, and at a pressure of 3000 or 4000 if the medium is oil, or otherwise expressed, fracture at superposed tensions of 1500 and 3000 or 4000, respectively. But if the external medium is Neoprene or lead or copper we have seen that tensile fracture occurs at superposed tensions of 5900, 15,500, and 24,500 kg/cm$^2$, respectively. There is a specific effect of hydrostatic pressure concealed in this comparison, but this effect does not obscure the main result. The surface conditions at glass are doubtless complicated, and at least in the case of water involve a chemical factor as well as a purely mechanical factor, since it is known that water is forced by pressure into the surface layers of glass. Brittle substances thus differ essentially from the substances which flow plastically before yielding in tension which have been made the subject of previous study. In the case of these ductile substances, fracture starts in the interior on the axis, where the tensile stress is a maximum, because of the stress redistribution produced by the necking itself, and always at values of the tensile stress materially higher than the ambient hydrostatic pressure. Fracture under these circumstances is again an energy-releasing process, because although work is done against the external pressure when a cavity appears in the inside, nevertheless the work received from the tensile stresses more than compensates because of their greater intensity. The fact that fracture starts at the inside explains the consistency of the results on fracture of ductile substances as contrasted with the capriciousness of the results for a brittle substance like glass.

From the point of view of energy release it is easy to understand the very much greater strength of glass to compressive stresses when supported by hydrostatic pressure than to tensile stresses. In the experiments above under a hydrostatic pressure of 27,000 or 28,000 kg/cm$^2$ the maximum superposed tensile stress supported was 24,500 kg/cm$^2$ (copper sheath), whereas a superposed compressive stress of 47,000 was observed. When fracture occurs under tensile stress combined with hydrostatic pressure two processes occur, both energy-releasing. The force producing the tension does work, and the pressure does work, because when tension is released the volume of the specimen decreases. But when fracture occurs under one-sided compressive stress, only one of the processes is energy-releasing. The force producing the compressive stress again does work during the fracture, but, since the volume increases on release of compressive stress, work is done against the pressure by the fracture.

From the point of view of energy release there is no particular significance as to whether the superposed tensile stress under which fracture occurs is greater or less than the hydrostatic pressure, that is, whether the tensile fracture occurs against a net tensile or compressive stress. For in either case the process of fracture may be resolved into two components, both energy-releasing. The significant thing is the lateral conditions, and the extent to which the surrounding hydrostatic pressure may follow through into any crack which appears, doing work in the process. In the case of ductile materials, when a crack appears on the surface the propagation is by a process of slip, which is not energy-releasing for the pressure, so that the fracture does not spread.

Both the strong dependence on the surface conditions in brittle fracture under hydrostatic pressure and the condition of energy release emphasize that there can be no single criterion of fracture in terms of stress and strain at a point only.

The actual experimental conditions differ from those of the idealized experiment in another important respect than the rigidity of the sheath, namely in the molecular structure of the specimen under test. In the idealized experiment, the material was thought of as structureless, and capable of indefinite and microscopically homogeneous distortion, as in the mathematical equations of elasticity theory. Actually, the material is composed of molecules, which have a certain impenetrability preventing indefinite approach, so that when strain is pushed too far individual molecules must be forced out of their positions by a series of discrete operations. It is doubtless something of this sort that is responsible for the remarkable separation of the glass tension specimens into thin disks, often, but not invariably, observed after simultaneous exposure of glass to pressure and tension. As tension increases, lateral contraction increases, and there must ultimately come a time when some of the molecules are brought into such close contact that they are forced out of their normal positions into neighboring spaces that have been opened by the simultaneous longitudinal extension. This process of molecular transfer is irreversible. When stress is released these displaced molecules act like internal wedges, distending the structure, which eventually fractures as in tension when stress is reduced far enough. A similar sort of thing has been found in other situations; fracture on release of stress would appear not to be uncommon. A somewhat similar state of affairs may be brought about in glass if foreign molecules are forced into the interior at the surface. Thus disking was almost always found in glass after a failure by the pinching-off effect in water as the transmitting medium.

The question has been much discussed of the great discrepancy between the experimental values of fracture stress and those calculated theoretically from various points of view. These considerations may have some bearing on this question. Strength in tension for a brittle substance we have seen to be an indeterminate matter, depending on the surface conditions. Under hydrostatic pressure breaking strengths in tension may be expected, with proper surface conditions, up to at least the magnitudes of the hydrostatic pressure. With pressures of the magnitude reached in these experiments, this is bringing the breaking stress up within reach of the theoretical values. In compression, dependence on the surface conditions is not to be anticipated, because the surface action that would take place at fracture is not energy releasing. Very large values of compressive strength can be reached, as for example 47,000 kg/cm$^2$ in Pyrex glass against some 2000 under ordinary conditions, or an increase of some 25-fold, again getting within sight of theoretical expectations.

It seems probable that the increases in compressive strength brought about by hydrostatic pressure are much greater for glass-like substances than for crystals. This is caused by the interlocking of the molecules in a glass, which is accentuated by pressure, making relative displacement more difficult. The effect is analogous to the very large increase of viscosity under pressure observed in some liquids with complicated molecules. In a crystal, on the other hand, the regular structure provides the possibility of internal slip. The plastic distortion of the crystal of $Al_2O_3$ is a case in point. The behavior of a crystal under compression is doubtless a highly specific property of the kind of crystal; I have never been able to produce measureable flow in crystalline quartz, although experiments on the collapsing of negative crystals show that some flow must be present.

Certain of the points emphasized here recall the point of view of Poncelet in a paper in Volume VI of Alexander's *Colloidal Chemistry*. Poncelet regards brittle fracture in tension as originating at the surface, and also emphasizes the role played by the "particulate" structure of ordinary matter as distinguished from the homogeneous isotropic structure assumed in the mathematical theory of elasticity.

Rowe and Gurney[8] have concluded from their experiments on the fracture of glass under "radial" pressure (my "pinching-off" effect) that Griffith's theory of fracture is essentially correct. According to this theory an actual specimen of glass is filled with minute crevices, with corners so sharp that under ordinary tensile load there may be stress concentrations at the corners by as much as a factor of 600. The breaking strength of geometrically perfect glass would thus be of the order of 6,000,000 p.s.i., which is of the order of the theoretical value, whereas the strength of an actual piece of glass is of the order of only 10,000 p.s.i. Under radial pressure fracture starts at the surface cracks. Rowe and Gurney have shown that assuming stress concentrations of 600, failure under radial pressure would be expected to occur at a pressure very approximately equal numerically to the ordinary breaking stress in tension, and this does in fact approximately agree with their experimental results. I think it must be admitted that qualitatively the Griffith point of view has considerable to recommend it. The strong dependence of fracture on surface conditions, for example, is consistent with its picture.

Qualitatively, Griffith's picture would lead to the expectation of pressure coefficients of tensile strength of completely sheathed brittle material of the order of magnitude of unity, and this also agrees with the experimental results above, and again indicates a real value in Griffith's point of view. It would not be easy to make a more precise calculation of the exact coefficient to be expected for completely sheathed specimens, however, because of the complicated nature of the surface conditions, and various other over-simplifications discussed in the next paragraph.

I question, however, whether other considerations do not play a vital role and whether the Griffith's picture is not much over-simplified. For instance, the surface conditions are in fact complicated, and involve chemical as well as purely mechanical factors, as shown by the very great effect of pressure media such as water or alcohol as contrasted with the oil used by Rowe and Gurney. Again, the ordinary concepts of stress and strain are carried down to a fineness of scale where there are stress concentrations of 600, whereas stress and strain are properly macroscopic concepts. At this scale of magnitude, the conditions of fracture are applied in the form that fracture occurs when the maximum tension reaches a critical value, irrespective of the other components of stress and independent of the strains. This condition of fracture does not seem to me to have any greater plausibility on the microscopic scale than on the macroscopic scale. The assumption of such a criterion amounts, among other things, to postulating that a pure hydrostatic pressure exerts no specific effect on the fracture properties of a substance. This is certainly not the case for ductile substances, as shown by my experiments on the enormous increase of the ductility of steel under pressure. The same thing is shown by the experiments of Rowe and Gurney themselves on plastics, which under their conditions of radial pressure tolerated an elongation before fracture eighteen times greater than under normal conditions. These results can properly be described as an increase of ductility under pressure. If plastic materials have their properties thus drastically altered by hydrostatic pressure it would seem to be only reasonable to expect some effect also on brittle substances. This is strongly suggested by the great increase of strength to simple compressive stresses found above for glass under high pressure.

---

[8] C. Gurney and P. W. Rowe, Second reference under 5.

The high tensile strength acquired by Carboloy under pressure is worthy of notice. The tensile stress at fracture is much higher than the surrounding hydrostatic pressure in the case of this material, so that the net stress at fracture is a tension. It is not necessary in this, or similar cases, therefore, to suppose that surface conditions play any role in the process, and the increase of strength may be a specific effect of pressure. The absolute value of the tensile strength, between 50,000 and 60,000 kg/cm$^2$, is not far from that which has been observed in steel, pulled to the limit of ductility and strain-hardened under pressure. A possible explanation of the effect in Carboloy is that it is associated with the cementing film of cobalt on the surface of each grain of carbide; the cobalt becomes ductile under pressure and strain-hardens in a way similar to steel.

A question that naturally presents itself in connection with these experiments is under what conditions does a substance lose its brittleness and become ductile under pressure? In order to deal with this question it is necessary to consider the concept of brittleness itself. It is not uncommon to speak of a brittle fracture, and in fact this usage has been followed in this paper. Such usage, however, employs a specialization and extension of meaning, because "brittle" properly does not refer to the process of fracture, but to what comes before the fracture. If the substance fractures before it receives permanent set, as shown by a permanent change of dimensions on release of stress, then the substance is brittle; if the substance fractures after it has received permanent set, then it is plastic or ductile. So far as the act of fracture itself is concerned it may be characterized as a shearing fracture or a tensile fracture, but not as a brittle fracture. When a brittle fracture is spoken of, what is meant is the fracture of a brittle substance. But strictly, from this point of view, all fractures might be spoken of as brittle, for a plastic substance which has been work-hardened to the limit and then fractures is, when at the point of fracture, a brittle substance. There is thus a considerable verbal element in the concept of brittle. Furthermore, the concept is not sharp, because if measurements are increased in sensitiveness plastic flow may be discovered before fracture where formerly with rougher measurements none appeared. In practise all that can be meant is fracture with no noticeable preliminary distortion.

In answer to our question as to what substances lose their brittleness under pressure, it may now be said that as a rough qualitative matter a crystalline substance, particularly if it is cubic, is more likely to become measureably plastic under pressure than an amorphous substance like glass. The observed plasticity of crystalline $Al_2O_3$ described in this paper lends plausibility to this statement. The limits are not sharp, however, and there is enormous variation in the numerical magnitude of the effect, as illustrated by the variation from soft steel to beryllium. In fact, even in the case of quartz crystal or glass an infinitesimal amount of flow must exist in principle, as shown by the spontaneous breakage of these materials under some circumstances a long time after exposure to high pressure.

---

# Addendum and Erratum: the Thermoluminescence and Conductivity of Phosphors*

[J. App. Phys. 17, 743 (1946)]

ROBERT C. HERMAN AND CHARLES F. MEYER
*Applied Physics Laboratory, Johns Hopkins University, Silver Spring, Maryland*

(Received November 21, 1946)

IN a recent paper with the above title,[1] the variations in the luminescence and conductivity of zinc sulphide type phosphors during infra-red illumination were discussed. The solutions given for the differential equations describing these processes are valid only in the case in which the retrapping of electrons is negligible. Recently, Klasens and Wise[2] derived the solution for the

---

* A portion of the work described in this paper has been supported by the Bureau of Ordnance, U. S. Navy, under contract NOrd-7386.

[1] R. C. Herman and C. F. Meyer, J. App. Phys. 17, 743 (1946).

[2] H. A. Klasens and M. E. Wise, Nature 158, 483 (1946).

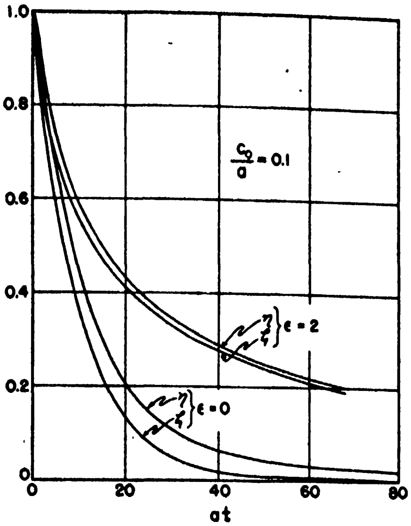


FIG. 1. The fraction of empty luminescent centers, $\eta$, and the fraction of trapped electrons, $\zeta$, *versus at* for $c_0/a = 0.1$ and $\epsilon = 0, 2$.

special case in which the cross section for capture of a conduction electron into a trap is equal to the cross section for capture into an empty luminescent center or hole. At the present time we are obtaining general solutions to this problem by means of a differential analyzer and wish to indicate some of the preliminary results.

The general solutions of Eq. (1) in reference 1 involve the following parameters[3]: $\epsilon$, the ratio of the cross sections for electron capture into empty traps and empty luminescent centers or holes and the parameter $c_0/a$, where $c_0 = \sigma I_0/h\nu$, and $a = AN_1$. In Fig. 1, the fraction of empty luminescent centers, $\eta$, and the fraction of trapped electrons, $\zeta$, are plotted *versus at* for $\epsilon = 0$ and 2 with the time $t$ in seconds. The corresponding variations of the luminescent intensity, $I$, proportional to $\xi\eta$, are plotted in Fig. 2. The behavior of $\xi$, which is proportional to the current $i$, can be obtained from Fig. 1 since $\xi = \eta - \zeta$.

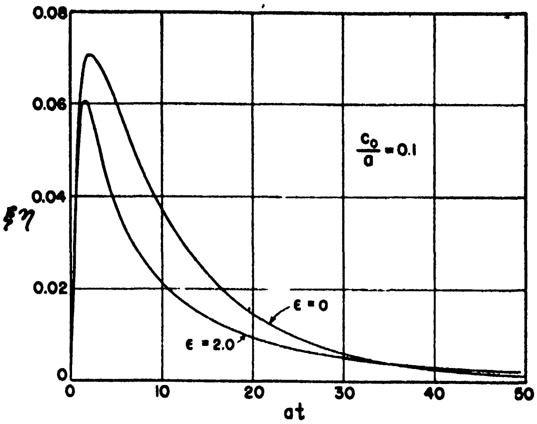


FIG. 2. The variations of $\xi\eta$, proportional to the luminescent intensity, *versus at* for $c_0/a = 0.1$ and $\epsilon = 0, 2$.

We take this opportunity to point out that Eq. (12) in our recent paper[1] should be written as follows:

$$\eta^{-1} = \exp(y - a/c_0) - (a/c_0)[Ei(-y) - Ei(-y_0)] \exp y$$

where $y = (a/c_0) \exp(-c_0 t)$, $y_0 = a/c_0$, $a = AN_1$, and $\eta = n_0/N_1$. This modifies Fig. 4 of the original paper in the manner shown for $\epsilon = 0$ in Figs. 1 and 2 of this note. The conclusions drawn from Fig. 5 remain unaltered.[1]

More detailed information on the effect of the retrapping of electrons on the luminescence and conductivity of phosphors during infra-red illumination as well as on the phenomenon of thermoluminescence will be published in the future.

We express our appreciation to Drs. G. C. Munro and D. T. Sigley of the Applied Physics Laboratory for their interest in this problem and thank the differential analyzer staff of the Moore School of Electrical Engineering, University of Pennsylvania for their cooperation.

---

[3] Note that $A = \sigma V$, where $\sigma$ is the cross section for capture into a hole, $V$ the velocity of a conduction electron; $I_0$ is the intensity of the incident infra-red light, and $N_1$ the number of electrons per cm$^3$ trapped originally.

# Letters to the Editor

## On the Action of Cutting Oils

G. P. Brewington
*Lawrence Institute of Technology, Detroit, Michigan*
December 5, 1946

ERNST and Merchant[1] have studied the efficiency (cutting ratio) of a number of cutting fluids on several metals and, for aluminum, have attempted to relate various physical properties of these fluids with observed cutting performance. They conclude ". . . that no relationship whatever appears to exist between the physical property and the cutting ratio." Even though the data given either in the paper referred to in reference 1 or in other sources are not as extensive as desired, it appears that some relationship does exist.

Those materials having a high cutting ratio appear to belong to that group of compounds having a sulfur ($-SH$) or chlorine in a low state of valence at the end of a short chain hydrocarbon. This would suggest that some of the higher valences of the sulfur and chlorine become active in the strong force fields known to exist at the newly formed surfaces of the work and chip, with the result that the molecule then adheres to the surface, thus reducing its activity and tendency to seize. The element fluorine exhibits only one valence but does form an analogous group of compounds similar to those of chlorine; it should be interesting, therefore, to compare the cutting ratios of these two fluids. The results obtained from cutting measurements on aluminum using a series of alcohols as cutting fluids may be due to a more common type of chemical action. If certain assumptions in the use of thermochemical data can be granted, the methanol-aluminum surface reaction must be sufficiently energetic to remove completely an aluminum atom, while the higher alcohols tend to attach themselves, through the $-OH$ radical, to the surface and form a monomolecular film. It would appear that an approach somewhat different from that used by Ernst and Merchant could be found in the surface energy studies of Bartell *et al.*[2]

The literature in the field of cutting oils is largely in the form of advertising and, while quite readily available, does not yield much quantitative information. Apparently innumerable experiments have been performed and reported only by word of mouth. Among such is a statement of high success associated with the use of a small stream of dry HCl gas playing on the cutting edge of a tool cutting steel. Should the HCl gas be as good as reported, and the cutting ratio as a function of cutting speed curve, as observed by Ernst and Merchant,[1] change materially, some indication as to the role of diffusion of the active agent in cutting phenomena might be obtained. Certainly the mobility of HCl gas would be much higher than that of larger organic molecules dissolved in liquids. It also should prove very interesting, if further data were available, to attempt to correlate the "cutting ratio" of various fluids with electron configurations of the active end of the molecule.

The cutting process viewed from any angle is extremely complicated and anyone who attempts to single out one phenomena for consideration is on dangerous ground. Assuming however, for this discussion, that tool failure at least depends on fatigue, which frequently can be associated with oscillation or standing wave phenomena, the varying force experienced by a tool tip must set up a standing wave pattern, possibly of ultrasonic frequencies, in the tool itself. This would suggest that the tool shape, particularly in carbide tools, should be such as to offer less opportunity for well-developed standing wave patterns to become established. One isolated unintentional experiment, reported verbally to the author, in which an "irregular chunk" of carbide brazed in a tool gave very good life, tends to suggest that other than the conventional form of tools might be further investigated.

The author wishes to thank Mr. H. A. Montgomery, president of the H. A. Montgomery Oil Company, for first suggesting that a problem, somewhat related to that discussed above, should be studied and also to Mr. W. Woijtowicz for many discussions of similar phenomena.

---

[1] "Surface Treatment of Metals," Symposium, Am. Soc. Metals (1940).

[2] Am. Chem. Soc. 56, 2205 (1934); J. Phys. Chem. 40, 881, 889, 895 (1936).

---

## Approximation for Absorption Displacement in X-Ray Diffraction by Highly Absorbing Cylindrical Samples

Zigmond W. Wilchinsky
*Esso Laboratories, Standard Oil Company (N. J.), Baton Rouge, Louisiana*
November 18, 1946

IN an article by Warren[1] treating displacement of diffracted x-rays caused by absorption in cylindrical samples, it was shown that for very high absorption, the displacement $\Delta S$ on the film is given by

$$\Delta S=(1+R/D)r\cos\theta\left\{\frac{\sin 2\theta-2\theta\cos 2\theta}{4\sin\theta+4\cos^2\theta\ln\tan\left(\frac{\pi}{4}-\frac{\theta}{2}\right)}\right\}, \quad (1)$$

where $R$ is the radius of the Debye camera used, $r$ is sample radius, and $D$ is the distance from the sample to the point

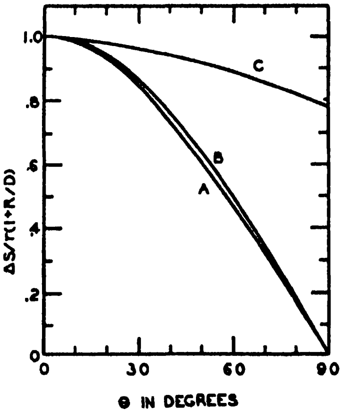


FIG. 1. Comparison of exact expression curve *A* with approximation curve *B*. Ratio of *A* to *B* is shown in *C*.

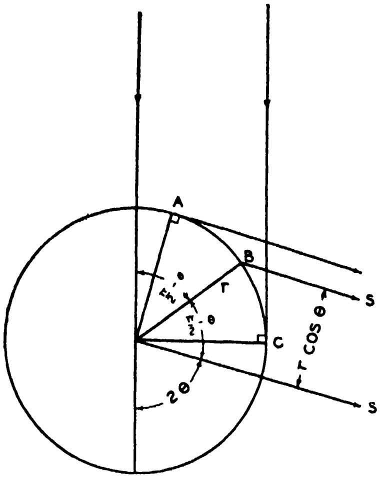


FIG. 2. Diffraction of a parallel x-ray beam from the surface of a highly absorbing cylindrical sample. By virtue of the above construction, *B*, the center of arc *ABC* lies on the bisector of the supplement of 2θ giving the above relationship among the angles.

from which the beam diverges. The purpose of this note is to point out that an approximation useful for general work is simply

$$\Delta S=(1+R/D)r\cos\theta. \quad (2)$$

This approximation is compared with Eq. (1) in Fig. 1. It can be noted that the approximation is best for small values of $\theta$. As $\theta$ increases, the magnitude of the departure increases at first, goes through a maximum, then decreases to zero at $\theta=90°$. The fractional departure however, increases steadily as can be gathered from curve *C* which is the ratio of (1) to (2), or simply the bracket in (1). At $\theta=90°$, curve *C* attains the value $\pi/4$.

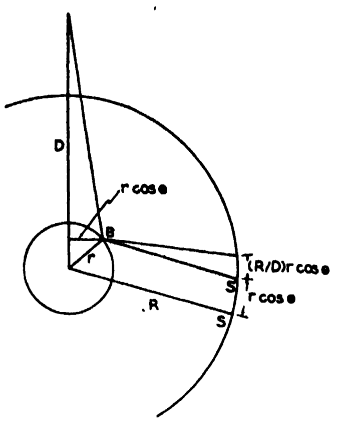


FIG. 3. Additional displacement caused by beam divergence.

A physical interpretation of the approximation (2) can be obtained with the aid of Fig. 2. First consider a parallel incident beam. For the right half of the sample, only the surface generated by moving *ABC* parallel to the sample axis is capable of contributing to the diffraction corresponding to $\theta$. *Assuming* that the center of gravity of the diffracted beam coincides with the ray diffracted from *B* the center of arc *ABC*, we find that *S'* is displaced parallel to the ideal position *S* by an amount $r\cos\theta$.

If the beam diverges from a point distant *D* from the sample (see Fig. 3), the angle of divergence of the ray to *B* is very nearly $(r/D)\cos\theta$. Therefore at the film, the ray diffracted from *B* is deviated by an amount $R(r/D)\cos\theta$, in addition to the quantity $r\cos\theta$ for the parallel case; the sum of the two effects is Eq. (2).

---

[1] B. E. Warren, J. App. Phys. 16, 614 (1945).

# Here and There

### New Appointments

J. H. Dillon, formerly with the Firestone Tire and Rubber Company, is now director of research of the Textile Research Institute, Inc., and of The Textile Foundation.

Frank M. Clark, insulation expert at the Pittsfield Works of the General Electric Company's Apparatus Department and assistant engineer of the Works Laboratory there, has been named technical consultant on insulation of the entire Apparatus Department.

W. Ewart Williams, English scientist now residing in the United States, has been appointed a consulting physicist on the research and engineering staff of Bausch and Lomb Optical Company.

### Awards

The Medal of Freedom, the nation's second highest decoration for civilian achievement in wartime, was awarded recently to two members of the staff of the National Bureau of Standards: Lauriston S. Taylor, chief of the X-Ray Section, and Fred L. Mohler, chief of the Atomic Physics Section. Mr. Taylor was also awarded the Bronze Star Medal.

### Electron Jubilee Celebrations

The fiftieth anniversary of the discovery of the electron by the British physicist, Sir Joseph Thomson, will occur this year. To mark this jubilee and to demonstrate the tremendous influence such an advance in pure physics may have on the life of the community, the Institute of Physics and the Physical Society are jointly arranging a series of meetings and other functions to take place on September 25 and 26, 1947, in London. A special exhibition which will remain open to the public for several weeks will be held at the Science Museum, South Kensington, and will show the development of the vast range of modern industrial equipment from its earliest experimental origins.

### Radio Engineering Show in March

The Annual Radio Engineering Show, a part of the 1947 National Convention of The Institute of Radio Engineers, will be held in Grand Central Palace, New York, March 3–6, 1947. The displays will be strictly of an engineering nature—transmitter equipment, instruments, component parts, and radio direction and location devices.

### Commemoration Program at Argonne National Laboratory

Major General Leslie R. Groves and Enrico Fermi were the principal speakers at the anniversary program held December 2 at the Museum of Science and Industry in Chicago, commemorating the initial operation of the first self-sustaining nuclear chain reaction.

### Research Physicists and Engineers Needed by NACA

The following statement was issued recently by the National Advisory Committee for Aeronautics:

"Physicists and engineers with research experience in problems related to aeronautics are needed by the National Advisory Committee for Aeronautics. Areas of research include the following: electronics, electrical measurements, servomechanisms, optics, interferometry, spectroscopy, electron or x-ray diffraction, mechanics of static or moving bodies, thermodynamics, aerodynamics, hydrodynamics, gas dynamics, metallurgy, lubrication, heat engines, mathematics, and other fields.

"The work of the NACA is to solve the fundamental problems of flight. This work is accomplished by means of basic research, consisting of experimental investigations and the development of theories needed for the prediction and explanation of aeronautical phenomena. The information obtained from this basic research is disseminated through publications to the field of aeronautics, where its application leads finally to the design and development of the nation's aircraft.

"This work is carried forward at three laboratories. The Langley Memorial Aeronautical Laboratory, Langley Field, Virginia, and the Ames Aeronautical Laboratory, Moffett Field, California, were constructed for, and are principally devoted to aerodynamic studies of the problems of flight and aircraft structures. Both are devoting a major portion of their work to fundamental problems of transonic and supersonic flight. Langley is also doing advanced research on the problems of guided missiles.

"The Aircraft Engine Research Laboratory, Cleveland, Ohio, is specifically designed for aircraft propulsion research. At the present time, the Cleveland facilities are primarily devoted to research investigations of the turbo jet, turbo propeller, ram jet, and rocket type power plants.

"Applications for positions or requests for further information should be addressed to the personnel officer of one of the laboratories."

## Calendar of Meetings

| | |
|---|---|
| **February** | |
| 15 | American Geophysical Union, Pasadena, California (Section of Hydrology) |
| 16–19 | American Institute of Chemical Engineers, Louisville, Kentucky (regional meeting) |
| 20–22 | Optical Society of America, New York, New York |
| 22 | American Mathematical Society, New York, New York |
| 24–25 | Inter-Society Color Council, New York, New York (Joint Meeting with the Technical Association of the Pulp and Paper Industry on the 25th) |
| 24–27 | Technical Association of the Pulp and Paper Industry, New York, New York |
| 24–28 | American Society for Testing Materials, Philadelphia, Pennsylvania |
| **March** | |
| 2–5 | American Society of Mechanical Engineers, Tulsa, Oklahoma |
| 3–6 | Institute of Radio Engineers, New York, New York |
| 17–21 | American Institute of Mining and Metallurgical Engineers, New York, New York |
| 22–27 | American Society for Metals, Oakland, California |
| 28 | Institute of Aeronautical Sciences, Cleveland, Ohio |
| **April** | |
| 4–5 | American Physical Society, Southeastern Section, Salisbury, North Carolina |
| 9–11 | Society of Automotive Engineers, New York, New York (Aeronautic Meeting) |
| 14–18 | American Chemical Society, Atlantic City, New Jersey |

Journal
of
Applied Physics

Volume 18, Number 3 March, 1947

# Ordnance Research Laboratory

By Eric A. Walker*
The Pennsylvania State College, State College, Pennsylvania

THE purpose of the Ordnance Research Laboratory at The Pennsylvania State College is to provide the Navy with an organization which can do research and development in an academic atmosphere. It differs from the Naval Ordnance Laboratory at White Oak, Maryland, and the Naval Research Laboratory in one important respect, namely, that this organization is completely under civilian control and has no Naval officers on its scientific or administrative staff. The organization of such laboratories as the Naval Ordnance Laboratory is too well known to demand further discussion at this time. It is recognized that such an organization cannot be perfect in all respects, and this new Laboratory represents an attempt to meet the problem of providing research for the Navy in a slightly different way.

Although the Laboratory was officially founded on January 1, 1945, it did not start work in its new building until October 1, 1945. However, the organization is much older than this. Many of the members of the staff were recruited from the Harvard Underwater Sound Laboratory, which itself was founded in 1941 and devoted most of its energies to submarine detection problems. Because of the caliber of some of the scientists, the Harvard Underwater Sound Laboratory was asked to undertake an ordnance problem which eventually grew into many and much larger problems, so that by 1943 the Ordnance Division of HUSL was firmly established. The major portion of this group was moved to the Ordnance Research Laboratory during the summer of 1945.

The Laboratory itself is housed in a two-story brick structure built especially for this purpose during the summer of 1945. This building affords approximately 30,000 square feet of space laid out in offices, laboratories, and shops in accordance with the needs. The Laboratory also operates two field groups, one at Newport, Rhode Island, and another at Fort Lauderdale, Florida.

FIG. 1. Exterior of Ordnance Research Laboratory. This Laboratory was designed by Paul Philippe Cret, well-known architect. It is laid out to be both functional and architecturally attractive.

---

* Director, Ordnance Research Laboratory.

FIG. 2. A view of the Black Moshannon Calibration Station for underwater sound devices. This calibration station is set on piles over a lake in a state forest. The general appearance is that of a Swiss chalet, and it was designed specifically to harmonize with its rural surroundings.

The one at Fort Lauderdale, Florida, will shortly be moved to a new laboratory structure at Key West. Both of these laboratories are modern buildings with office space, laboratory, and shops, but their work primarily is that of field testing the finished product or of obtaining fundamental determinations on which working designs can be based. At each of the field stations there exists a small core of four or five men who are on a permanent basis. The remainder of the staff is composed of scientists who are sent to the field stations for periods of from a few days to several months in order to test their designs.

The relationship between the Laboratory and the College is relatively simple. The Laboratory is established as a department of the School of Engineering in exactly the same way as any other department, such as Aeronautical, Civil, Electrical Engineering, etc. All members of the Laboratory staff are regular employees of the College. The scientists are given the academic titles of Professor, including the usual Associate and Assistant Professorships, or of Research Assistant, which in the usual academic scale corresponds to Instructor. Technicians hired as specialists in a particular field are usually classed as Assistants. The remainder of the staff, including machinists, clerks, and secretaries, carries appropriate designation. The direct line of responsibility between the Laboratory and the remainder of the College is through the Director and the Dean of the School of Engineering.

The relationship with the Navy is a much more tenuous one than that existing with the College, since there is only one Naval officer attached to the Laboratory. In fact, the Navy, by a contract between the College and the Bureau of Ordnance, sets aside a certain fund for the Laboratory to use in doing research and development on underwater ordnance. The Naval Ordnance Officer is expected to be cognizant of all the activities of the Laboratory and must approve certain business transactions which, because of their financial magnitude, seem to require prior Navy approval. The scientific tie between the Laboratory and the Navy is almost placed on a personal basis. The Navy does not write a directive on which the Laboratory must act, but most research programs undertaken by the Laboratory originate with the scientific staff of the Laboratory and are then submitted to the appropriate Naval officers for their approval. Because liaison between Naval officers and the scientific staff is a rather close and cordial one, usually there are a number of discussions preceding any formal proposal, and to date things are so well prepared that no proposal has been made which has not been accepted. This method of problem approval is a valuable one from the Laboratory's viewpoint because it serves as an automatic check on the work of the Laboratory and insures against scientists going off on unprecedented problems, while allowing the staff to judge the worth of the investigation and, to a great extent, to follow their own interests.

The internal organization of the Laboratory is not particularly novel, but it must stress certain factors which have to do with the product of the Laboratory, namely, scientific research. The Laboratory is headed by a Director and several Assistant Directors, but the weekly review of the working operations comes under the cognizance of a Council. This Council consists of the Directors of the Laboratory, including the Business Manager and Personnel Director, the leaders of all the sections, and the project engineers in charge of the major projects of the Laboratory. The scientific staff is organized both horizontally and vertically. There are five major scientific sections: Acoustics, Electronics, Applied Mechanics, Hydro- and Aerodynamics, and Mathematics. These may be further subdivided.

For instance, the Electronics Section is divided into a research division, a development division, and a production division, which undertake the particular functions indicated.

The Applied Mechanics Section is divided in a different manner, since it must fill a service function for the remainder of the Laboratory. It includes a design division which is expected to carry on all mechanical design through the blueprint stage, and the shops, which are expected to produce the mechanisms. This section also employs a number of mechanical designers or engineers who are expected to carry on any research projects which fall into this classification and to serve as representatives of this section in a manner which will be outlined later.

The Acoustics Section has research, development, and construction functions, and also test responsibilities which include the test of any acoustical device which is produced in this Laboratory or a device on which tests may be requested by other organizations.

The Hydro- and Aerodynamics Section is currently being organized. It will consist of two major divisions, the first concerned with the mathematical aspects of aero- and hydrodynamics with special emphasis on body shapes, propellers, and cavitation. The other division will be concerned with operations for the test of these devices either at sea or in the water tunnel.

The Mathematics Section, because of the peculiar nature of its work, has little or no formal organization.

The work of the Laboratory is usually concentrated in a small number of major projects. These projects are almost entirely development work, and therefore they come under the cognizance of a project engineer. In assembling a staff for any particular project, the project engineer may feel that he needs the services of engineers or scientists skilled in particular scientific fields, so he may request the proper section leader to assign to him a scientist who is familiar with the type of problem which he expects to meet. Conversely, the section leader, knowing a project is to be undertaken, may decide that this particular project has a number of features in which his group will be interested. He may then request representation and assign to the group a scientist who will pass on all of the designs which come in his field and who will keep the section leader informed of the progress of the work.

The work in a project usually divides itself into a construction phase which is undertaken at the main laboratory and a test phase which is undertaken at one of the field stations. One of the inviolable principles of operation is that the group which tests a device must be the same group which designed and supervised its assembly at the home laboratory. This serves to preserve continuity of knowledge which exists in unwritten form and which may be of considerable importance in the design and final operation of the device. Some difficulties arise from time to time when an engineer is asked to serve both the project engineer and the section leader. This matter is usually resolved by a ruling that in the field the project engineer is in complete charge, while in the laboratory he must confer with the section leaders.

Every attempt is made to insure complete dissemination of all information about all projects. This is possible because they cover a narrow range of scientific activity. First, a weekly organization report is issued. The material is collected as painlessly as possible by the group secretaries, who visit the various individuals at their work places and ask for their weekly reports. In it each scientific worker in the Laboratory may have his say. Frankness is encouraged, and because the organization report

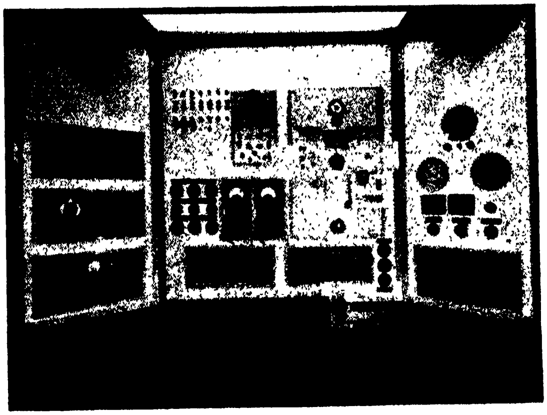


FIG. 3. Front view, Black Moshannon console. A close-up view of the console from which all of the underwater acoustic measurements are made. This panel is so laid out that all measurements can be made from this point and recorded automatically without the operator's leaving the control position.

FIG. 4. This is a view of one of the general shops. This particular shop is devoted to light machine work for the construction of models and test equipment.

never leaves the confines of the Laboratory, no editing is encouraged or attempted. The notes published in this organization report may be lengthy or brief, depending on the scientist's mood and on the amount of work he has done during the past week. At times these reports may be deadly reading, but they give every member of the Laboratory a chance to be heard on any subject, scientific, administrative, or otherwise.

The second method is to encourage the frequent writing of internal memoranda. Each member of the staff is told that when he reaches a suitable stopping point in his work, he should write a memorandum addressed to his immediate superior and sent to anyone else in the Laboratory who he feels might be interested in his work. As is expected, these memoranda provoke considerable comment and much constructive criticism.

The third method is by conferences. These are almost never scheduled on a calendar basis but are assembled on short notice by the project engineers when they feel that some clarification is needed.

The reports of the Laboratory take two forms: a monthly report which tells of the progress of all research programs, and reports on a single program which are issued when a suitable tying-off point is reached. Ultimately, the product of the Laboratory will be gadgets or weapons, together with complete sets of prints by which they may be constructed, and the maintenance and instruction manuals which are necessary for their use. Completion reports will include an evaluation of the device which sums up the matter by either saying that the gadget is no good and should be forgotten, or that it is useful and should go to the Fleet for training purposes.

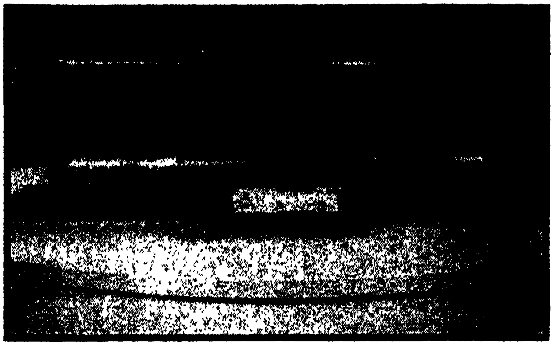


FIG. 5. Scale model of the 48-inch water tunnel which will be built at Penn State in the near future. The large section of the tunnel has a diameter of 12 feet and it is over 100 feet long. The picture shows a full-size railway box car to the same scale.

An attempt is made to provide every useful mechanical aid for the scientists, since they, at the present time, are by far the scarcest commodity. Computers are provided to make routine calculations and to plot curves. Draftsmen are provided to put sketches into publishable form, and a library and staff keep the master files and all available useful information. Because these files become quite voluminous it has been found desirable to keep a microfilm file of all documents which are six months old or older. Most of these documents are on the general theme of underwater ordnance and include reports of this and other laboratories, as well as technical accounts of the ordnance used by foreign governments.

In the administration of any such organization, certain aids and methods of operation appear which are peculiar to the personnel employed and which may work in a laboratory of one size and nature and not in another. Experience at ORL has shown that it is desirable to provide almost complete segregation between the business and the scientific functions of the Laboratory. No scientist should be asked or even allowed to waste his time on matters which belong in a purchasing office. He should not be allowed to hire personnel because he may make commitments which cannot be honored. He should not be asked to waste his time securing tickets or hotel reservations if it can better be done by someone attached to the Business Office. One other peculiarity arises from the fact that in a research organization very often only one item of a kind is made. Many scientists feel that such an item can best be designed "on the lathe." It is quite obvious that no machine shop of any size can be run in this manner, so at ORL the stipulation is made that all machine shop work is done either to blueprint or to approved sketches. This still leaves the possibility that in some cases the scientist's contention may be correct, and it is not worth the time to make detailed drawings or elaborate sketches before construction is started. For this reason, a small and well-equipped machine shop is provided in which any scientist can do his own work. For the safety of both the scientist and the machines, an expert machinist is always in attendance who may actually do the job or, if necessary, guide the scientist while the job is being done. Such an arrangement, although a rather elaborate one, seems to remove many causes of irritation and friction.

Every attempt is made to get publications into the scientific press, and for this purpose an Editorial Office exists and may be called upon to give assistance if the author wishes or requires it.

# Here and There

### New Appointments

G. Howard Carragan has been appointed head of the Department of Physics at Rensselaer Polytechnic Institute. He succeeds Robert A. Patterson, who recently left to serve as an assistant director of Brookhaven National Laboratory, atomic research center at Camp Upton, Long Island.

### Walter Evans Receives Army Award

The War Department has announced the award of the Army Certificate of Appreciation to Walter Evans, vice president in charge of all radio activities for the Westinghouse Electric Corporation, for "his contribution to the Signal Corps in connection with the development and production of radio and radar equipment during World War II."

### Duane Roller Receives Oersted Medal

The American Association of Physics Teachers has awarded the Oersted Medal to Duane Roller, professor and head of the Department of Physics of Wabash College, Crawfordsville, Indiana. The award is made annually "for notable contributions to the teaching of physics."

Through the *American Journal of Physics*, which he has edited since it was founded in 1933, Dr. Roller had made probably the most notable contribution to the teaching of physics at the college and university levels since the subject was introduced into our educational system. If the *Journal* were his sole contribution (which it is not), Dr. Roller would richly deserve this signal honor.

He attended Culver Military Academy and did his undergraduate work at the University of Oklahoma. He also received the Master's degree at Oklahoma and taught there for thirteen years partly before and partly after his work for the Doctorate at California Institute of Technology, which he received in 1929. Subsequently he was a research associate at Columbia and an associate professor at Hunter College. During the first world war he was a pilot and during the recent war he was Chief Technical Aide of the National Defense Research Committee.

Former recipients of the Oersted Medal, which has been awarded every year since 1936, include W. S. Franklin of Lehigh, M.I.T., and Rollings; R. A. Millikan of Columbia, Chicago and Caltech; Henry Crew of Northwestern, and G. W. Stewart of Iowa. Last year's award was to R. L. Edwards of Miami University.

The formal award was made at the January 31 session of the annual meeting of the American Association of Physics Teachers, which occurred at Columbia University. In his address of acceptance Dr. Roller discussed one of his chief professional interests, physical terminology, another field in which he has made notable contributions, both to the teaching of physics and to the science itself.

### Chicago I.R.E. Conference

The second Chicago Conference of the Institute of Radio Engineers will be held on Saturday, April 19, at Northwestern Technological Institute. It will feature an all-day series of technical sessions and discussions on the practical side of electronic engineering, with emphasis on applied electronics.

### The Bulletin of Mathematical Biophysics

Following is the table of contents of the March 1947 issue of *The Bulletin of Mathematical Biophysics:*

**Suggestions for a Mathematical Biophysics of Some Psychoses**—N. RASHEVSKY

**A Problem in Mathematical Biophysics of Interaction of Two or More Individuals Which May Be of Interest in Mathematical Sociology**—N. RASHEVSKY

**Mathematic Theory of Motivation Interactions of Two Individuals: I.—**
**ANATOL RAPOPORT**
**The Mechanism of the Middle Ear: I. The Two Piston Problem—**
**MARTINUS H. M. ESSER**

The University of Chicago Press, Chicago, Illinois, Volume 9, Number 1.

### New Physics Laboratory at Carnegie Tech

The Buhl Foundation of Pittsburgh has made a grant of $300,000 to Carnegie Institute of Technology for the partial support of a plan to build a 200-million-volt synchrocyclotron and nuclear physics laboratory. A condition of the grant provides that at least $250,000 shall be subscribed by others for the same purpose. Dr. Charles F. Lewis, Director of The Buhl Foundation, said of the grant, "It is confidently expected that this program will notably strengthen Pittsburgh's position as one of the nation's outstanding centers in scientific research."

### Institute of Navigation

The Institute of Navigation held its eastern regional meeting on February 13 and 14 in New York City. The Institute was founded during the war by members of the Armed Services and by civilians as a non-profit scientific and technical society devoted to the modern science of navigation in the air and on the sea. The membership is composed chiefly of astronomers, mathematicians, navigators, aviators, physicists, and engineers. The president of the Institute is Commander Gordon A. Atwater, Director of the Hayden Planetarium, New York City.

### Frontiers in Chemistry Lectures

The sixth annual series of Frontiers in Chemistry lectures at Western Reserve University, Cleveland, Ohio, will occur in February, March, and April. Among the twelve scientists to appear in the series are Dr. Glenn T. Seaborg, Dr. Harold C. Urey, and Dr. K. C. D. Hickman. The lectures will be divided into two sections. The first will deal with recent advances in physical and inorganic chemistry, and the second with modern theories of organic chemistry. The course is designed chiefly for graduate chemists and physicists.

### Physics Research at Northwestern

In January it was announced that Northwestern University would begin immediately a $500,000 three-year program of fundamental research in physics that is expected to lead to new and improved communication devices. Financed by the Navy, the program will be directed by Professors Walter S. Huxford and Robert J. Cashman of the Department of Physics. It is an extension of wartime research in which they developed a new invisible-ray telephone and many new types of photoelectric cells used in the phone and in various Army and Navy secret weapons.

### Hugh L. Dryden Elected to International Committee

Hugh L. Dryden, associate director of the National Bureau of Standards, was elected a U. S. representative to the International Committee at the Sixth International Congress for Applied Mechanics held at the Sorbonne, Paris, September 1946. The Congress brought together for the first time since the war those scientists from each major country interested in theoretical and experimental work on the elastic and plastic properties of materials and structures, fluid mechanics, vibration, sound, friction and lubrication, thermodynamics, combustion, and heat transfer. The International Committee, consisting of outstanding scientists from member nations in the field of applied mechanics, is the managing body of the Congress, which is concerned with the international cooperation, advancement and standardization of the sciences of applied mechanics.

### General Electric Fellowships at Case School

Case School of Applied Science will offer fifty fellowships to high school teachers of physics for a six-weeks program of study during the summer of 1947. Recognizing the fact that industrial research and progress stem largely from a knowledge of physics, The General Electric Company has provided these fellowships for high school and preparatory school teachers of physics. The program is designed to acquaint teachers with recent scientific developments. The fellowships include all tuition fees, room and board, and travel expenses. High school and preparatory school teachers of physics from the following states are eligible to apply: Ohio, Michigan, Western Pennsylvania, West Virginia, Kentucky, Indiana, Illinois, Wisconsin, and Maryland.

### Research in Mineral Flotation at MIT

Armour and Company has established a fund of $12,000 for a graduate research program in the fundamentals of mineral flotation at the Massachusetts Institute of Technology. This program, which provides for grants-in-aid for several assistantships and fellowships, will concentrate on the operation of cationic collectors, particularly the organic compounds known as amines and amine salts.

### Harvey L. Curtis Retires from Bureau of Standards

Dr. Harvey L. Curtis, physicist at the National Bureau of Standards since 1907, has retired. His major scientific achievement has been in the field of absolute electrical measurements. He has written the standard textbook on the subject, and his conclusions will form the basis for United States proposals in establishing a fixed relationship between the absolute and international electrical measurement units at the next International Electrical Congress. During World War II he assumed responsibility for interior ballistics research at the Bureau.

### Journal of Mathematics and Physics

The Journal of Mathematics and Physics is published by the Massachusetts Institute of Technology in annual volumes of four numbers. Papers which appear in this journal deal with mathematical problems from the fields of engineering and applied physics and with mathematical methods of interest for the applications.

Individual members of the American Institute of Physics may subscribe to the Journal at the special rate of $3 per year. The regular subscription price is $4 per year. Subscription orders and other correspondence may be sent to the Journal of Mathematics and Physics, Massachusetts Institute of Technology, Boston 39, Massachusetts.

### Reinforced Concrete Research at Bureau of Standards

The American Iron and Steel Institute is supporting a research program on reinforced concrete now under way at the National Bureau of Standards. The prime objectives are improved design of concrete reinforcing bars and engineering data that will enable engineers and builders to make more effective use of reinforced concrete.

### Awards

Two members of the staff of the Naval Research Laboratory recently received the Distinguished Civilian Service Award from the Secretary of the Navy for their contributions to the war effort: George R. Irwin, director of the ballistics section, and Wayne C. Hall, head of the electrical division.

The War Department has awarded the Bronze Star Medal to Harold O. Wyckoff, associate physicist in the X-Ray Section, National Bureau of Standards. Dr. Wyckoff served in the European Theater as assistant chief of the Operational Research Section of the Ninth Air Force.

### Radio Engineering Meeting

On May 3 in Cambridge, Massachusetts, there will be an all-day New England Radio Engineering Meeting under the sponsorship of the North Atlantic Region of the Institute of Radio Engineers. All persons interested in radio and electronic engineering are cordially invited to attend. There will be six technical sessions, none held concurrently, exhibits by the leading manufacturers in New England, a luncheon and a banquet. The entire program will be held at the Continental Hotel in Cambridge. A complete program announcement will be made later.

### German Physical Society

A conference of German physicists was held October 4–6, 1946, at Göttingen, Germany, in the British Zone of Occupation. About three hundred physicists were present. A German Physical Society was organized for the British Zone. Following are the titles of papers presented: "Theory of the diffraction of electro-magnetic waves on a circular disk," "Measurements on the slowing-down of neutrons," "On neutron processes in beryllium, aluminum and lithium," "Experiments on photographing by means of neutrons," "Changes of sensitivity on light counter tubes," "On a method for the determination of energy and energy distribution of ionizing particles," "On the age of the world," "On the isotope displacement effect in the *E I*-Spectrum," "Experimental detection of the von Schmidt headwave in optics," "Collision discharge in highest pressure mercury lamps," "The accommodation time of the electron temperature in a stationary electric discharge," "Determination of the altitude of the sodium layer shining in the night sky," "Oxide layers on metals," "The optical and electrical qualities of synthetic Greenockite," "Density of sublimed salt layers," "Thermal dissociation in the radiation black body cavity," "Deformation of electron clouds by the collision of gas molecules," "Intuitive models for the theory of super-conductivity," "Superconductivity and the periodic system of elements," "Electron theory of superconductivity," "Sublimation, condensation, saturation pressure and critical point in a model experiment," "Does the lattice constant change for very small particles?" "Optical analyses on electrolytic solutions. Measurements on absorption coefficients," "Thermal effects of gliding friction," "Equation of state of water at high densities," "Exact viscosity measurements up to 2000 atm." "Hydrogen bonding, resonance and the formation of molecular chains," "Remarks on the statistical theory of turbulence."

Dr. Max von Laue was elected Chairman of the Society. The question of journals was discussed but no action was taken.

The following note at the end of the program is significant. "This program may at the same time be regarded as a ticket. . . . Ration cards, including those for potatoes, are necessary."

### Necrology

George A. Scott, assistant professor of physics at the University of Pittsburgh, passed away suddenly following a heart attack in January.

---

## PROCEEDINGS OF THE ELECTRON MICROSCOPE SOCIETY OF AMERICA

THE annual meeting of the Electron Microscope Society of America was held at the Mellon Institute in Pittsburgh, Pennsylvania, December 5–7.

Titles and abstracts of the papers presented are given below:

**1. Trends in X-Ray Diffraction Work.** DAVID HARKER, *General Electric Company, Schenectady, New York.*

**2. Trends in the Use of the Electron Microscope in the Study of Fine Structure.** JAMES HILLIER, *Radio Corporation of America, Princeton, New Jersey.*

**3. The Application of Electron Diffraction Methods in the Study of Fine Structure.** ROBERT HEIDENREICH, *Bell Telephone Laboratories, Murray Hill, New Jersey.*

**4. First-Order Theory of the Three-Electrode Electron Gun with Plane Electrodes.** S. G. ELLIS, *University of Toronto, Canada.*—On the simplifying assumption of linear potential variations within the electron gun it is possible to deduce the position of the image of the cathode, the magnification and the angular aperture of the beam external to the gun. The results are compared with those of Johannson. The area of the specimen illuminated in the electron microscope is discussed. It is shown that in general an electron gun can be operated either with or without the production of a cross-over.

**5. Some Observations on the Shadow Cast Formvar Replica Technique.** BEATRICE M. DEACON, S. G. ELLIS, AND W. G. CROSS, *University of Toronto, Canada.*—Formvar replicas of polished glass surfaces containing sleeks were examined in the electron microscope both with and without shadow casting. In another experiment several replicas from the same region of a glass surface were examined. The nature of the replicas and the limitations of the method are discussed.

**6. Electron Microscopy of Colloid Systems.** JOHN TURKEVICH, *Princeton University* AND JAMES HILLIER, *Radio Corporation of America, Princeton, New Jersey.*—Typical colloidal systems such as gold, silver, platinum, alumina, ferric oxide, tungsten oxide, vanadium pentoxide, cadmium sulfide, arsenic sulfide, etc., were examined under high resolution with the electron microscope. The form of the colloidal particle and its size were determined as a function of the variables used in the preparation. A method will be described for obtaining gold particles of very uniform size and suitable for calibration standards. The behavior of particles of alumina on steam treatment will be illustrated with electron microphotographs.

**7. Electron Microscope Examination of the Microphysical Properties of the Polymer Cuprene.** JOHN H. L. WATSON, *Shawinigan Chemicals Limited, Shawinigan Falls, Quebec.*—The microphysical structure of cuprene is described from electron micrographs and suggestions are made from them concerning the mechanism of the formation and catalysis of this material. Copper is detected in the electron diffraction pattern. Cuprene samples prepared by polymerization of acetylene in the presence of finely divided cuprous oxide are shown as hollow fibers characterized by both longitudinal and transverse structure and a very thin skin. The striking appearance of this type of cuprene serves to identify it in clogging deposits taken from acetylene lines. This is illustrated by a deposit taken from a reactor used in the process of hydrogenation of acetylene to ethylene. "Cuprene" specimens formed by alpha-ray bombardment of acetylene gas are shown as round particles joined by short necks about 500 to 1000A in length. An oxidized sample has round particles of mean diameter 4900A and unoxidized sample of mean diameter 3800A. The alpha-ray cuprene is quite unlike the catalyzed solid material in appearance and in addition possesses characteristics of a very viscous liquid in the electron micrographs.

**8. An Effect of Electron Bombardment upon Carbon Black.** JOHN H. L. WATSON, *Shawinigan Chemicals Limited, Shawinigan Falls, Quebec.*—A significant change is reported in mean particle size and shape characteristics of carbon black due to specimen contamination while under examination in electron microscopes. The effect is described for a number of well-known commercial blacks and examples are given in micrograph, graphical and tabular form to illustrate expected variations in mean particle size with continued bombardment. At normal focusing intensities, the particle size usually will have changed more than 5 percent in 30, or sometimes in as little as 15 seconds. The change is detected with other materials but is more pronounced with carbon black. Chromium shadow cast specimens seem to change to a less degree than uncoated samples. The contamination is explained as arising from the well-known polymerization and condensation processes which occur when organic vapors and gases are bombarded by charged particles. Suggestions are made for minimizing the effect.

**9. The Electron Microscope Examination of Greases.** S. G. ELLIS, *University of Toronto, Canada.*—A method is described for examining the soap component of lubricating greases with the electron microscope. The soap component is mounted on Formvar films from a dispersion in ether. The specimen is shadow cast by the method of Williams and Wyckoff. The deduction of the shape of the soap particles is discussed.

**10. Artifacts Produced by Poor Resolution.** L. H. WILLISFORD, *Goodyear Tire and Rubber Company, Akron, Ohio.*—A micrograph of round objects may show the following two opposite conditions: (1) the individual objects of a group will have a tie or rod connection with each adjacent individual of that group and (2) isolated individual objects will have smooth profiles with no evidence of any portion of a tie or rod. This condition of rod connection is caused by imperfect resolution. A simple set-up has been devised for casting shadows which simulate continuous changes from good to poor resolution. When two disks under conditions of poor resolution approach each other, the space between them remains clear until the distance is just less than the limit of resolution, at which distance a misty connection appears. With closer and closer approach, the connection darkens; and before the disks are in contact the connection has the appearance of reality. When slightly overlapped, the disks have a pronounced dumbbell effect. This artifact rod has been confirmed by destroying the resolution of a standard micro-tessar lens; in fact, the "rod" effect is quite universal, and an audience participation demonstration will be given which will duplicate the clues of poor resolution to be looked for in regular and stereo micrographs of round objects.

**Some structural patterns in biological systems.** FRANCIS O. SCHMITT, *Massachusetts Institute of Technology, Cambridge, Massachusetts.*

**11. Remarks on the Borderline of Physics, Chemistry, and Biology.** L. MARTON.—A dual conception of life phenomena is proposed in a manner somewhat similar to the dual conception of wave mechanics. The probability of "creation of order from order" depends on the size of the elementary unit of a chemical compound and on the *non-periodic* complexity of its components. "Life" is therefore not restricted to the group of so called "living organisms," the probability function has a finite value for the simplest chemical compounds, although its value may be immeasurably small. For large chemical compounds, which are considered often on the border line between chemistry and biology, it depends on the experiment whether the "chemical" or the "life" character will be predominant. These considerations necessitate also a critical survey of the molecular concept. A more statistical definition of the molecule is attempted and experiments are proposed to prove the conclusions of the paper.

**12. Applications of the Phase Microscope in Electron Microscopy.** OSCAR W. RICHARDS, *American Optical Company, Buffalo, New York.*—The Spencer Phase Microscope controls light passing through it so that invisible phase differences arising from optical path differences in the specimen are converted into visible amplitude or intensity differences. In phase microscopy the contrast in the image may be increased, decreased or reversed, and either increased or decreased. The instrument and its use will be described. The Phase Microscope, within its limits of resolution, is useful for the examination of transparent specimens (e.g., replicas). The relation of phase and electron microscopy will be discussed.

**13. Preparation of Tissue for High Speed Sectioning.** MARY C. SCHUSTER AND CLIFFERD E. GRAY, *Interchemical Corporation, New York, New York.*—The standard techniques of tissue preparation for light microscopy produce distortions that become apparent with the increased resolution of the electron microscope. An adaptation of standard technique is made in the following manner: the tissue is washed free of soluble proteins using isotonic saline solution; fixation is started at very great dilution, and the concentration of fixative is increased slowly to the desired level; when fixation is complete, dehydration and embedding are performed with similar care. Preparation of various tissues with various reagents is discussed and illustrated. It is believed that the procedures described will add to the effectiveness of the joint techniques of high speed sectioning and electron microscopy.

**14. Electron Microscope Studies on the Structure of Chlorophyll "Crystals," "Crystalline Alpha Carotene," "Crystalline Beta Carotene" and "Crystalline Vitamin A."** EDITH A. ROBERTS, *Vassar College, Poughkeepsie, New York.*—Electron microscope studies of the above show that they are composed of structural units less than 100 ängströms, some of units close to 10 ängströms. These units appear to be protoplasmic entities comparable to those present in plastids rather than "crystals."

**15. Further Electron Microscope Studies of the Structure of the Starch Grain.** MILDRED D. SOUTHWICK, *Vassar College, Poughkeepsie, New York.*—Electron microscope studies of the structural organization of starch grains, either fresh or commercially prepared from the seeds, roots, stems, or leaves from different genera of plants belonging to the *Bryophyta*, *Pteridophyta*, and *Spermatophyta* indicate the same essential morphology of all starch grains. All show the presence of structural units, comparable to those in the chloroplast—the larger units, plastidules, and the smaller units, plastid granules, as well as the units of a still smaller size which would approach the smallest functional protoplasmic units known.

**16. The Mounting of Bacteria for Electron Microscope Examination.** JAMES HILLIER AND R. F. BAKER, *Radio Corporation of America, Princeton, New Jersey.*—In the case of some bacteria grown on solid media it is possible to remove a young culture from the medium intact. In other cases a modification of the light microscopists impression technique is found to be very satisfactory.

**17. A Study, with the High Voltage Electron Microscope, of the Endospore and the Life Cycle of Bacillus Mycoides.** GEORGE KNAYSI, *Cornell University, Ithaca, New York* AND R. F. BAKER AND JAMES HILLIER, *Radio Corporation of America, Princeton, New Jersey.*—A study of strain $C_2$ of *bacillus mycoides* at 150 kv shows that the endospores obtained from old agar-slant cultures present considerable differences in appearance. Most common are those which are uniformly opaque, those which show a black outline and a shrunken content, and those with transparent lateral areas and thick folds. The exine is thick and surrounded by a thin, elastic pellicle. The intine can sometimes be seen. Germination is initiated by an increased transparency to the electrons and by lateral bulging. This is followed by cracking of the exine, the two halves of which usually remain attached but occasionally are totally severed. In ordinary media, the germ cell and subsequent generations are semi-transparent but show no differentiation in their protoplasm until the fatty inclusions which usually appear before sporulation are formed. We were unable to observe these granules at 50 kv. The forespore appears homogeneous. The mature spore often shows a dark outline and a slightly shrunken protoplasm within the mother cell. High voltage is of distinct advantage in the study of this problem.

**18. Demonstration, with the Electron Microscope, of Nucleus-like Bodies in Cells of *Bacillus Mycoides* Grown in Nitrogen-Free Media.** GEORGE KNAYSI, *Cornell University, Ithaca, New York* AND R. F. BAKER, *Radio Corporation of America, Princeton, New Jersey.*—The endospore of *bacillus mycoides* contains a relatively considerable quantity of ribonucleic acid which is an excellent source of nitrogen but a very poor source of energy. In ordinary nutrient media, this material is also readily formed by the vegetative cell and is diffuse throughout the protoplasm in an apparently combined form. It is the reason why such cells stain homogeneously and are opaque to electrons. When the endospore germinates in a medium without a nitrogen source, the ribonucleic acid is used up and the subsequent generations of vegetative cells become poorly stainable and transparent to electrons even at 30 kv. In such cells one observes one to several opaque bodies which appear to divide and which are finally enclosed in the forespore; this behavior indicates their nuclear nature. There is also a variable number of smaller bodies of unknown significance. The best results were obtained with a medium consisting of 0.2 g of glucose+0.2 g of Na acetate+100 ml of distilled water.

**19. Observation on Actin and Myosin Extracted from Muscle.** M. A. Jakus and C. E. Hall, *Massachusetts Institute of Technology Cambridge, Massachusetts.*—Actin has been extracted from striated muscle according to the method of Straub. The addition of neutral salts or hydrogen ions to an aqueous solution of actin brings about a reversible transformation of the globular protein into a fibrous form. Electron micrographs of actin at different stages of linear aggregation will be shown. A water-soluble muscle component designated as "myosin" by Szent-Györgyi has also been prepared and examined. Micrographs of this protein will be presented and compared with those of myosin extracted in alkaline salt solution after the method of Greenstein and Edsall.

**20. Size Distribution of Tobacco Mosaic in the Early Stages of Infection.** Gerald Oster, *The Rockefeller Institute for Medical Research, Princeton, New Jersey.*—A method is described whereby the contents of the hair cells from leaves of Turkish tobacco plants diseased with tobacco mosaic virus may be examined by means of the electron microscope. It was found that 68 percent of the rod-like particles present in the hair cells of plants infected for 25 days are about 280 mu in length. The contents of hair cells infected with the virus were also examined less than one day after innoculation, and particles less than 280 mu in length were found to be present. The size distribution of virus preparations subjected to sonic treatment was determined. The virus activity of the sample is proportional to the number of particles 280 mu in length that are present.

**21. Studies on the Newcastle Disease Virus (California Strain).** A. R. Taylor, *Duke University School of Medicine, Durham, North Carolina.*—The infectious agent responsible for Newcastle disease of fowls has been concentrated and purified by ultracentrifugation. Studies of the concentrates have been made by means of electron micrography, analytical ultracentrifugation, and chemical analysis. The virus is a sperm-shaped particle with a head of about 70-mu width and 180-mu length and a thin tail of about 500-mu length. Well-defined internal structure was observed in the head-piece. The virus consists of protein and lipid in association with a small amount of mucleic acid. The results of elementary and component analyses will be given in detail.

**22. An Electron Microscope Study of the Structures of Some Virus Protein Crystals.** R. W. G. Wyckoff, *National Insitute of Health, Bethesda, Maryland.*—This is a discussion, illustrated with electron micrographs, of the kinds of regularity of particle arrangement obtained when suspensions of several crystallizable macro molecules are desiccated in the presence and in the absence of salts. These kinds of order are compared with that observed by electron microscopy in preformed crystals of these substances.

**23. Immunochemical Aspects of Sonic Treated Tobacco Mosaic Virus.** Saul Malkiel, *The Rockefeller Institute for Medical Research, Princeton, New Jersey.*—In preparing tobacco mosaic virus from the juice of infected plants, Sigurgeirsson and Stanley found small non-infectious particles about ⅓ of the usual length of the tobacco mosaic virus particles in the supernatant liquids after centrifugation of the normal virus particles. Preparations of this small particle were compared quantitatively with the $280\times15$-mu sized rod by means of the precipitin reaction. More antibody was precipitated by the rod of normal size. A solution of freshly prepared tobacco mosaic virus was subjected to sonic vibrations in order to yield small particles. By fractional centrifugation, the normal sized tobacco mosaic virus particles were removed, leaving a supernatant liquid containing a preponderance of smaller particles as checked by distribution measurements of electron micrographs. In order to investigate and compare more fully the immunochemical behavior of these particles with normal tobacco mosaic virus, the stoichiometric relationship of antigen to antibody in specific precipitates was determined by means of the quantitative precipitin reaction. Both heterlogous and homologous antisera were used for this investigation. These findings and the possible theoretical considerations are presented.

**24. Specimen Preparation for Electron Micrographs and Diffraction Patterns.** Leland L. Antes, *University of Texas, Austin, Texas.*—Several simple and effective methods of dispersing fine powders which have a tendency to agglomerate when in suspension are outlined. As applied to such materials as carbon black, metallic oxide pigments and catalysts, these methods may be roughly classified into dry smear and surface film techniques. The results of these methods are compared with others which are generally used. Methods of preparing water-soluble salts for diffraction patterns are discussed, and the advantages of electron diffraction methods in determining crystal structure are pointed out. A means of directly comparing the intensities of diffraction pattern spots or rings is suggested.

**25. Alterations and Additional Equipment for an R.C.A. Console Electron Microscope.** Henry C. Froula, *Armour Research Foundation, Chicago, Illinois.*—Alterations and additional equipment were made for an R.C.A. console electron microscope to simplify maintenance and operation and to improve performance. The items involved include: (1) controls for more positive gun-alignment, (2) equipment for testing the vacuum system for leaks, (3) panel control of illumination, (4) calibration of intensity and focus controls, (5) functionally improved control dials, (6) cassette markings for more certain identification of exposures, and (7) simple holder for storage of specimens. The usefulness of these items will be discussed.

**26. Spherical Aberration of Compound Magnetic Lenses.** L. MARTON AND K. BOL, *Stanford University, Palo Alto, California.*

**27. Further Studies on the Magnetic Electron Microscope Objective.** J. HILLIER AND E. G. RAMBERG, *Radio Corporation of America, Princeton, New Jersey.*—Fresnel diffraction fringes observed in through-focus series are shown to be a sensitive criterion for the existence of asymmetries in the electron microscope objective. It is shown that sufficiently accurate symmetry in the objective field can be attained by empirical compensation. The method of compensating the objective is described in detail and some of the results in which resolving powers of 10A were obtained are shown. Further experiments with a new high contrast objective are described.

**28. Bias Focused Electron Gun.** JOHN H. REISNER, *Radio Corporation of America, Camden, New Jersey.*—History of the gun—A qualitative description of the action of the gun; field plots to be given; effect of circuits on the operation; physical parameters affecting spot sizes and intensities; desirable uses of the gun; effect on alignment; effect of vacuum; filament shapes; constructional materials.

**29. Photographic Plates for Use in Electron Microscopy and Electron Diffraction.** ROBERT G. PICARD AND W. F SWANN, *Radio Corporation of America, Camden, New Jersey.*—A summary of work carried out jointly by Eastman Kodak and R.C.A., to provide a plate especially suited to the requirements of electron microscopy. After many attempts, an emulsion having long scale, medium contrast, fine grain and adequate speed has been developed. Special, practically grainless plates admirably suited for electron diffraction and subsequent microphotometry have also been tested and will be described.

**30. Calculated Magnitude of Fresnel Diffraction in Shadow Casting.** C. E. HALL, *Massachusetts Institute of Technology, Cambridge, Massachusetts.*—The intensity curve of the diffraction pattern near the edge of a shadow cast by an evaporated beam is calculated for a Maxwellian distribution of velocities in the beam. Effects other than diffraction which may contribute to unsharpness of the edge are neglected. The results indicate that when heights of the order of 100A are to be measured accurately by the shadow-casting technique, the width of the shadow edge may be appreciable in comparison to the total length of the shadow.

**31. Electron Microscope Goniometry.** A. F. KIRKPATRICK, EVELYN GAGNON, AND T. G. RECHOW, *Stamford Research Laboratories, American Cyanamid Company, Stamford, Connecticut.*—The frequent occurrence of crystals in electron micrographs presents the problem of their identification. The silhouette angles observed on the photographic plate are functions of the true interfacial angles of these crystals. The interfacial angles of a suspected compound, available in the literature, can be used to calculate the angles that would be formed by the orthographic projection of the crystal edges upon the plane of the photographic plate. A comparison of the calculated angles with those measured on the electron micrograph serves to identify the crystals. Thus a measurement of a physical constant can be made with the electron microscope. The calculations can be made by geometrical construction with the use of stereographic projections. Examples of the use of electron microscope crystallography in the identification of crystals will be presented.

**32. Preparation and Uses of Silica Replicas in Electron Microscopy.** C. H. GEROULD, *The Dow Chemical Company, Midland, Michigan.*—A detailed description of the preparation of silica replicas and substrates and their many and varied uses is presented with photographs and electron micrographs. The question of "up and down" in stereo pictures of silica replicas is definitely settled by adding a metal oxide smoke (MgO) to the polystyrene replica before the silica evaporation. A successful powder dispersing technique consists of dispersing the powder in an ethyl cellulose lacquer and depositing upon a silica substrate. A method of preparing silica replicas of specimens which cannot be subjected to the temperatures and pressures of the ordinary technique is described. It consists of applying a polystyrene lacquer to the specimen surface in place of the conventional molding. Several other techniques are also included.

**33. Procedure for Preparing Latex Samples for Electron Micrographs.** WALTER E. BROWN, *The B. F. Goodrich Company.*—It is difficult to obtain good electron micrographs of latices consisting of soft particles such as natural and synthetic rubbers because the particles tend to flatten and merge. These particles can be hardened by bromination to give good micrographs. For GR-S and Hycar OR 15 this treatment results in a swelling of about 9 percent of the diameter. There is no other apparent change that might cause difficulty in interpreting the results. The reliability of the method is shown by: (1) comparison with a vulcanized GR-S latex; (2) by combining two latices of different particle sizes and obtaining the size distribution for the mixture; and (3) observing a particle in random orientation. The complete technique for making such samples is discussed. It is especially important that all of the materials used be very clean and this involves the use of freshly distilled water and solvents. Generally, it is necessary to wash the sample after drying.

**34. A Review of Some Recent Advances in Electron Microscopy Microtomy.** E. FULLAM, *General Electric Research Laboratories, Schenectady, New York.*

**35. Round Table Discussion of Cross Sectioning for the Electron Microscope—Invited Speakers and Discussion from the Floor.**

**36. Round Table Discussion on Particle Size Problems in Electron Microscopy—Invited Speakers and Discussion from the Floor.** MARY C. SCHUSTER, *Interchemical Cooperation, New York, New York,* Moderator.

# Contributed Original Research

## Theory of Radar Reflection from Wires or Thin Metallic Strips*

J. H. VAN VLECK,** F. BLOCH,† AND M. HAMERMESH††
*Radio Research Laboratory, Harvard University, Cambridge, Massachusetts*

(Received September 3, 1946)

Knowledge of the radar response of wires or thin metallic strips, as a function of their length and thickness, and of the radar frequency is important in the design of reflectors for radar. In view of the difficulty of this theoretical problem and the necessity of making approximations, as well as the dearth of adequate experimental data, two independent procedures for solution are presented. Detailed quantitative results are obtained for the angular dependence of the cross section, and also for the mean cross section, of randomly-oriented wires or, more generally, of metallic strips, which behave electromagnetically like cylindrical wires of a certain "equivalent radius." When expressed in terms of a unit of area equal to the square of the wave-length, these cross sections depend on the dimensions of the wire only through the two ratios

$$\frac{2l}{a}=\frac{\text{length of wire}}{\text{equivalent radius of wire}},\quad \frac{2l}{\lambda}=\frac{\text{length of wire}}{\text{wave-length}}.$$

The mean cross section is shown to take on maximum values when $4l/\lambda$ is slightly less than an integer ($n=1$, 2, etc.). The shift of these "resonances" from integral values depends on the ratio $2l/a$, becoming greater as $2l/a$ decreases. The value of $\bar{\sigma}/\lambda^2$ at resonance increases slowly with the order $n$ of the resonance; it depends only very slightly on the ratio $2l/a$, increasing as $2l/a$ decreases. For values of $4l/\lambda$ away from resonance, $\bar{\sigma}/\lambda^2$ decreases rapidly, reaching minimum values near $4l/\lambda=3/2$, 5/2, etc. The value of $\bar{\sigma}/\lambda^2$ at these minima is strongly dependent on $2l/a$, increasing as $2l/a$ decreases. Also as $4l/\lambda$ increases, the heights of the minima increase and approach the height of the resonance peaks. A brief comparison with preliminary experimental results is given.

### I. DESCRIPTIVE SURVEY

*INTRODUCTION:* In the use of radar, an electromagnetic wave is sent out from a transmitter; the wave impinges on a target and some of the energy is scattered back in the direction of the radar. There it is received and presented on a screen. The strength of the return signal (echo) will depend, among other factors, on the scattering properties of the target. These can be described completely, for our purposes, by introducing the concept of "radar cross section" $\sigma$ defined as $4\pi$ times the quotient of power returned per unit solid angle in the direction of the radar to the power density incident upon the target.

The customary targets for radar observation are aircraft, ships, and topographical features. Since these objects have not generally been constructed from the point of view of their use as radar reflectors, one finds that they are rather inefficient in this respect—one can obtain radar responses comparable to those of an aircraft or ship by using properly designed reflectors of much smaller dimensions. These reflectors can perform several functions.

They may be used to present echoes to a radar for purposes of identification of a target, or for marking a geographical location, thus replacing complicated beacons.

For military purposes they can be used to present "false" echoes to a radar operator for purposes of deception or confusion.

Useful types of reflectors for these purposes are flat sheets, spheres, corner reflectors, and metallic wires or strips. Reflectors for use by aircraft require special consideration as regards weight, size, and ease of handling. The most convenient type of reflector in this case is the bundle of strips, known as window, which, when dropped from the aircraft, will produce the required radar signal. Though strips are most convenient for

---

* The work reported in this paper was done at the Radio Research Laboratory, Harvard University under contract with Division 15 of the National Defense Research Committee.

** Now at Harvard University, Cambridge, Massachusetts.

† Now at Stanford University, Palo Alto, California.

†† Now at New York University, University Heights, New York, New York.

practical use, we shall refer to the reflectors as "wires," and assume that they are circular cylinders. This assumption involves no loss of generality since, as shown in another paper by F. Bloch and M. Hamermesh,[1] a thin antenna of arbitrary cross-sectional shape has the same electrical characteristics as a circular cylinder of the same length and having an equivalent radius $a$. For a flat strip the equivalent radius is one-fourth the width of the strip.

In designing these reflectors it is of importance to know the radar response of the wire as a function of its length and thickness, and the wave-length used by the radar. There will be an optimum length for producing echoes at a given radar wave-length; the band of wave-lengths to which the wire responds strongly will depend on its thickness. Wires cut for one wave-length may produce undesirable responses at other wave-lengths (effect of "harmonics").

The calculation of the radar response of wires is also of interest from the point of view of general antenna theory. The problem presents most of the pertinent features of the receiving or transmitting antenna; at the same time it does not involve the complication presented by the gap at the center in such instruments. Thus it represents our best opportunity for comparing theoretical predictions with experiment.

The problem of radar reflection from a wire can be formulated as follows: A transmitter emits a linearly polarized wave which impinges upon a wire, the direction of incidence being at an angle $\theta$ to the axis of the wire. The wave scattered back along this same direction is picked up by a receiver ordinarily polarized parallel to the transmitter. We wish to calculate the power returned to this receiver as a function of the angle of incidence $\theta$, the direction of polarization, the wave-length, and the length and radius of the wire.

*Notation:* We shall use the following symbols:

$\lambda$ = wave-length.

$\beta = 2\pi/\lambda$; $\omega = 2\pi c/\lambda$; $c$ = velocity of light.

$2l$ = length of wire.

$a$ = equivalent radius of wire.

$E_0$ = amplitude of the electric field of the incident wave.

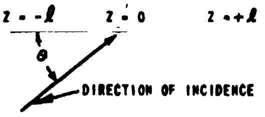


FIG. 1.

$\varphi$ = angle between the electric field of the incident wave and the plane of Fig. 1.

$q = \beta \cos\theta$.

$I(z)$ = current in the wire at $z$, the coordinate giving position along the wire (Fig. 1).

log = logarithm to base $e$.

The cross section $\sigma$ is defined as:

$$\sigma = 4\pi \frac{\text{Power reflected backwards (per unit solid angle)}}{\text{Power density incident upon the target}}$$

$\bar{\sigma}$ is the "average cross section" which will be observed when the signal is reflected from a bundle of randomly oriented wires. It can be computed from $\sigma$ by averaging:

$$\bar{\sigma} = (1/4\pi) \int\int \sigma \sin\theta d\theta d\varphi.$$

$\sigma(\theta) = (1/2\pi) \int \sigma d\varphi$ is the average return signal for fixed direction of incidence and random polarization. The average power returned to a rotating dipole antenna from a wire in fixed orientation would be given by $\sigma(\theta)$.

*Methods of Solution:* The first step in the solution of our problem is to calculate the current induced in the wire. (This is essentially the problem of the receiving antenna.)

A simple method, which has been used in similar problems is that of Siegel and Labus.[2] This method starts from a reasonable assumption concerning the form of the expression for the current, and evaluates the numerical magnitudes in this expression by applying the requirement of conservation of energy (e.m.f. method). Rigoristically speaking, it is not a method for deducing the solution to the antenna problem. A second, more satisfactory procedure, is to try to satisfy the Maxwell equations and to fulfill the boundary conditions at the surface of the wire, using a method of successive approximations. The tangential component of the total electric field must vanish on the surface of the wire (assumed to be a perfect conductor). This

---

[1] F. Bloch and M. Hamermesh, Report 411–TM125 of the Radio Research Laboratory (NDRC Division 15).

[2] E. Siegel and J. Labus, Hoch: tech. u. Elek: akus. 43, 166 (1934).

total field consists of the incident field plus the field produced by the induced current in the wire. For sufficiently thin wires the current at the ends may be set equal to zero. The mathematical equivalent of these statements is the integral equation[3]

$$\int_{-l}^{+l} d\xi I(\xi)\frac{e^{-i\beta r}}{r}=\frac{i\omega \cos\varphi E_0 e^{iqz}}{\beta^2 \sin\theta} + A_1 \cos\beta z + A_2 \sin\beta z \quad (1)$$

with

$$r=[(z-\xi)^2+a^2]^{\frac{1}{2}}, \quad I(\pm l)=0.$$

These equations assume that the current is located at the center of the wire rather than residing on its surface, but it is well known in antenna theory that the resulting error is not serious (the reason being essentially that any infinitely long cylindrically symmetric charge distribution gives the same exterior potential as though concentrated at the center).

King and Harrison[3] solved the integral Eq. (1) for the receiving antenna by a method of successive approximations due to E. Hallén.[4] In its original form, the Hallén method suffered from the defect that the convergence of the successive approximations becomes poorer as $l/\lambda$ increases. Even for $l=\lambda$, a reasonably accurate solution would require the use of complicated higher approximations.

This difficulty, however, is avoided by following procedures introduced by Miss Gray,[5] and especially by King and Middleton,[6] who start from the same integral equation but who choose different procedures of successive approximation. The Gray procedure was intended primarily for the transmitting antenna but, as explained more fully in the last part of the appendix, it is more suitable for the receiving antenna. It leads to difficulties when the wire is so very long that it contains a large number of wave-lengths (a case which does not usually arise in practice). The most refined scheme of approximation is that recently published by King and Middleton.

We have obtained solutions of the problem by two methods:

(A) A combination of a judicious guess based on the solution of King and Harrison (with some corrections à la Gray or King and Middleton), and the application of the principle of conservation of energy.

(B) An approximate solution of the integral equation for the receiving antenna. The procedure which we employed is one which we deemed at the time a modification of the Gray method, since it resembled the latter rather more than that of Hallén. After our investigations were completed, however, the paper of King and Middleton was published. It turns out that actually our scheme for approximating the integral equation involves the same kind of attack as that used by King and Middleton in the antenna problem, though the superficial language is rather different. The point is discussed more fully at the end of the appendix. The mathematical details of the approximate solutions by methods (A) and (B) will be described in Sections II and III, respectively.

In both methods (A) and (B) the current along the wire is assumed to be of the form

$$I(z)=C_1 \cos qz + C_2 \cos\beta z + C_3 \sin qz + C_4 \sin\beta z, \quad (2)$$

where $C_1$, $C_2$, $C_3$, $C_4$ are functions of the angles $\theta$, $\varphi$ and of $l$, $\lambda$, $a$. We shall refer to the terms in $\sin qz$ and $\sin\beta z$ (or $\cos qz$ and $\cos\beta z$) as the forced and resonant parts, respectively, of the solution, inasmuch as the phase of the impressed field along the wire is $qz$, and inasmuch as the amplitude $C_2$ (or $C_4$) is abnormally large, because of resonance, if $\beta l/\pi$ is a half-integer (or integer). The cosine and sine terms will be characterized as the even and odd members, respectively. In method (A) the values of $C_1$, $C_2$, $C_3$, $C_4$, are determined by the conservation of energy and a supplemental assumption concerning the form of the solution away from resonance. In (B) one calculates the $C_1$, $C_2$, $C_3$, $C_4$ in a more dynamical fashion from the integral equation. In (A) it is

---

[3] R. King and C. Harrison, Proc. I.R.E. 31, 548 (1943); and 32, 18 and 35 (1944).

[4] E. Hallén, Nova Acta (Uppsala), Series 4, 11, 1 (1938).

[5] M. C. Gray, J. App. Phys. 15, 61 (1944).

[6] R. King and D. Middleton, Quart. App. Math. 3, 302 (1946). For an interesting critical comparison of the various schemes which have been proposed for approximate solution of the integral equation for the transmitting antenna, see D. Middleton and R. King, J. App. Phys. 17, 273 (1946). Our approximation is, we believe, comparable with what these authors call "first-order King-Middleton," as we do not essay the refined but rather lengthy calculation of the second order terms included in the articles of Bouwkamp and of King and Middleton.

assumed that

$$\frac{C_1}{C_2}=-\frac{\cos\beta l}{\cos ql};\quad \frac{C_3}{C_4}=-\frac{\sin\beta l}{\sin ql},\tag{3}$$

so that the boundary conditions $I(\pm l)=0$ are automatically satisfied. In (B) the ratios $C_1/C_2$, $C_3/C_4$ have somewhat different values than (3), especially near resonance. At first sight this fact may seem to imply that in (B) the boundary conditions are not satisfied, but this is not really the case, as in satisfying the latter in (B) certain extra terms are included which we have not written down in the expression (2) for $I(z)$. It is consistent to carry these higher order terms only in satisfying the boundary conditions,[7] but not in computing the radiation, since most of the radiation comes from parts of the wire not near the ends, whereas the correction terms are abnormally important at the ends, especially in the interesting resonance region where $\cos\beta l$ or $\sin\beta l$ is nearly zero. In (A) the amplitude of one of the forced terms drops to zero at resonance, since (3) shows that $C_1$ and $C_3$ vanish with $\cos\beta l$ and $\sin\beta l$, respectively. Actually, this is not correct, as the forced vibration is unaffected by resonance rather than suppressed by it. In consequence of the fact that (B) uses a more refined current expression than does (A) in satisfying the boundary condition, (B) is able to yield various effects and details which (A) is incapable of describing.

*Chu's Theory:* A simple and ingeneous, though crude, method for obtaining expressions for the radar response, was used by Chu (unpublished). (We shall not give the details of this method since, as shown in Sec. II, the results can be obtained as a special case of Method A.)

For resonant wires (length a multiple of $\lambda/2$) Chu proceeds as follows: The current is assumed to be a pure sine function vanishing at the ends of the wire. Then the magnitude of the current is determined by applying the law of conservation of energy (as in our method (A), which differs from Chu in the choice of current form). The expression for $\bar{\sigma}$ at resonance is

$$\bar{\sigma}=\frac{3\lambda^2}{16\pi}\,\frac{4\pi^2 l/\lambda-\frac{3}{2}+3\log(8\pi l\gamma/\lambda)-\log 2}{[\log(8\pi l\gamma/\lambda)]^2},\tag{4}$$

where $\gamma$ denotes Euler's constant 1.78. The variation of cross section with angle of incidence, i.e., $\sigma(\theta)$, may be obtained from the diagrams on page 788 of Terman's Handbook. One must, however, take the fourth power of the radial coordinate shown in his plots, as the cross section involves power rather than voltage, and as there is an extra square because on back-scattering the angular factor occurs both in incidence and reflection.

The other case treated by Chu is that of long wires. Here only forced terms are considered (i.e., $C_2=C_4=0$ in Eq. (2)) and he obtains

$$\sigma=\frac{4\pi l^2\sin^2\theta(\sin x/x)^2}{(\pi/2)^2+(\log\{\lambda/\gamma\pi a\sin\theta\})^2}\cos^4\varphi,\tag{5}$$

with the abbreviation $x=(4\pi l/\lambda)\cos\theta$. Averaging over all $\theta$ and $\varphi$, Chu finds

$$\bar{\sigma}=\tfrac{3}{8}l\lambda\frac{\pi/2-\lambda/8\pi l}{(\pi/2)^2+(\log\{\lambda/\gamma\pi a\})^2}.\tag{5a}$$

Chu's method has the advantage of extreme simplicity, but it suffers from several defects:

(a) It gives no information on the response in the intermediate region between the peaks corresponding to the resonant lengths, and the valleys corresponding more to the case of long wires.

(b) It does not predict any shifts in the resonance maxima from the values $\lambda=4l/m$, where $m$ is an integer. Actually, the experimentally observed maxima are somewhat on the long wave-length side.

(c) At resonance it neglects the effect of the forced oscillation, which is not negligible for broadside response or at the higher maxima (large values of $m$).

(d) It makes an error in the values of the height of the resonance peaks because the resonance shifts (effect $b$) alter the damping constant, and hence the amplitude.

(e) The heights of the valleys are furnished by the Chu theory only if one assumes that away

---

[7] The current can be made to vanish at the end even in the first approximation if one iterates in such a way that at each stage the solution at the end of the wire is subtracted from that at the middle. This is done, for instance, in the papers of Hallén and of King and Middleton. Such a difference in procedure from ours is, however, a rather superficial one. It amounts, so to speak, to merely a different system of book-keeping for the various terms which necessarily appear as the iteration progresses.

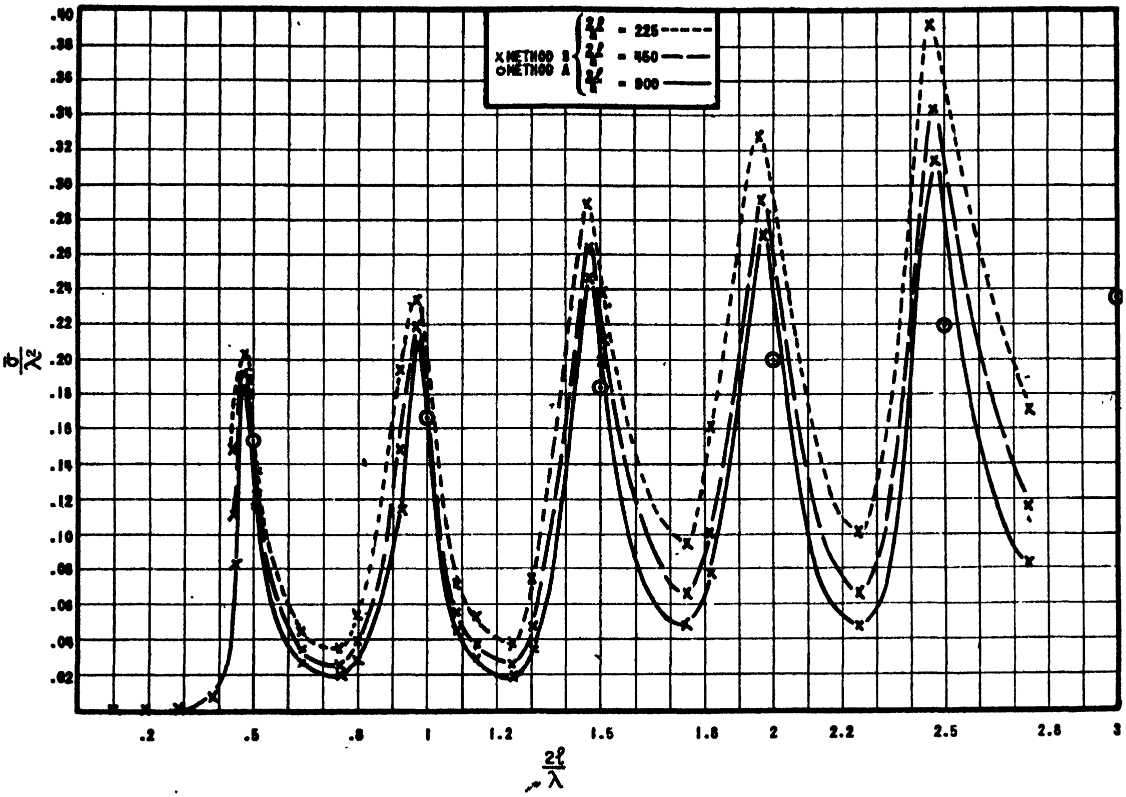


FIG. 2. Radar response of wires: average cross section.

from resonance practically all the response is caused by the forced rather than resonant terms, so that one can use his result (5a) for infinitely long wires. Actually, for wires of practical length the effect of the resonant members cannot be neglected even in the valleys.

*Relative Merits of Methods* (*A*) *and* (*B*)*:* The results of Chu may be roughly characterized as a zero-order theory, which suffers from the defects mentioned in items (a)–(e) inclusive. Method (A) gives formulas for all wave-lengths, thus removing difficulty (a), but it is open to the objections (b–c–d). It is more than just an interpolation scheme between the various Chu formulas as it does include the effect of the resonant terms away from the maxima, and so is not subject to the drawback (e). Hence (A) usually agrees quite closely with (B) in the valleys, but near the peaks (A) gives results which practically coincide with those given by the Chu resonance formula. Quite irrespective of the additional refinement connected with (b), (c), (d), method (B) has the advantage that it yields formulas that are much more tractable than those supplied by (A), as it gives closed expressions for the cross section averaged over all angles. Hence most of our comparison with experiment will be based on (B) and, because of the labor involved, only a few points will be computed by means of (A).

## Results

(1) *Variation of the Averaged Cross Section with Wave-Length:* It is perhaps well first to take a bird's-eye view of the response curves furnished by the theory for the cross section averaged over all angles, i.e., for a random distribution of wires. This is provided by Figs. 2 and 3. The curves are those furnished by method (B). Each curve of Fig. 2 refers to a different value of the ratio $2l/a$ of the wire length to radius. In Fig. 2, the cross section is plotted out to $l/\lambda=1.25$ (five peaks), while in Fig. 3 the curve is extended to the 20th peak for the special case $2l/a=900$. The points actually calculated by method (B) are indicated by crosses in Fig. 2, and it is seen that, except for a few selected instances, most of the points

beyond the second or third maximum have been calculated only for the peaks and valleys, i.e., for resonance and antiresonance. No monumental accuracy should be attached to the precise values of the coordinates in the transition region where the curves slope very steeply and cannot be determined precisely without calculating a very large number of points. In the additional region covered by Fig. 3, only the peaks and valleys have been calculated, except that the shape of the curve has been computed in some detail for the 19th peak in order to examine whether there are any important changes in the band width at the higher resonances. It is seen from Fig. 3 that, measured in frequency, the band width increases slowly as one passes to the higher resonances, and is over 3 times as great for the 19th as for the first peak. On the other hand, the percentage change in frequency required to detune from resonance is obviously much smaller at the higher peaks, simply because the frequency is higher. As one would expect, the peaks are sharper, the smaller the radius of the wire in comparison to its length. No attempt has been made to calculate a detailed curve by method (A). Instead, the points calculated with this method are indicated by circles in Figs. 2 and 3. As can be seen from the curves, the two methods agree quite well away from resonance. The circles at $2l/\lambda=\frac{1}{2}$, 1, $1\frac{1}{2}$, $2\cdots$ are the peak maxima for method (A), and have practically the same values as those given by the simple resonance theory of Chu.

It should be noted that the percentage difference between resonance and antiresonance diminishes as the length of the wire, measured in multiples of wave-length, increases. For instance, Fig. 3 shows that the first peak is ten times as high as the first valley, whereas the ninth peak has only three times the height of the ninth valley. For infinitely long wires the difference between the peaks and valleys would disappear entirely.

As we go from one valley to another the height of the valley floor does not increase uniformly, but rather increases in an alternating fashion. The first and second valleys, for instance, are approximately equal in elevation. Similar remarks apply to the third and fourth, which are both considerably higher than the first and second. This result is yielded by both (A) and (B), and is in accord with the experimental data.

(2) *Positions of the Maxima:* Both experimentally and theoretically (method (B)), the resonance maxima come at somewhat longer wave-lengths than in the usual approximate formula $4l/\lambda=m$. The displacements are greater the larger the radius of the wire; for an infinitely thin wire the resonance would be at exactly $4l/\lambda=m$. With finite radii, the expression for the resonance maxima furnished by method (B) is approximately

$$\begin{pmatrix}\cot\beta l\\ -\tan\beta l\end{pmatrix}=\tfrac{1}{4}\pi[2\log(\lambda/\pi a)+\tfrac{1}{2}\log(2\pi l/\lambda)-1.87]^{-1}. \quad (6)$$

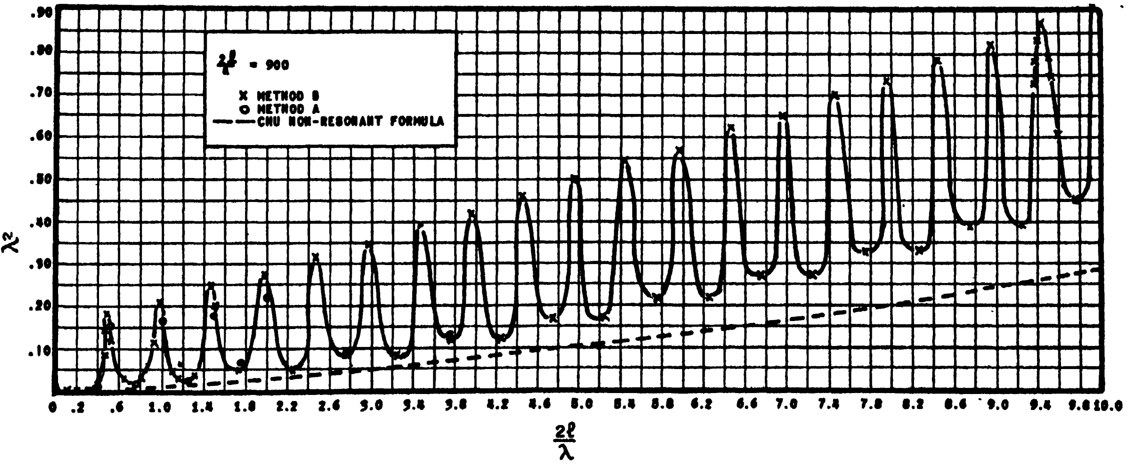


FIG. 3. Response of wires: average cross section.

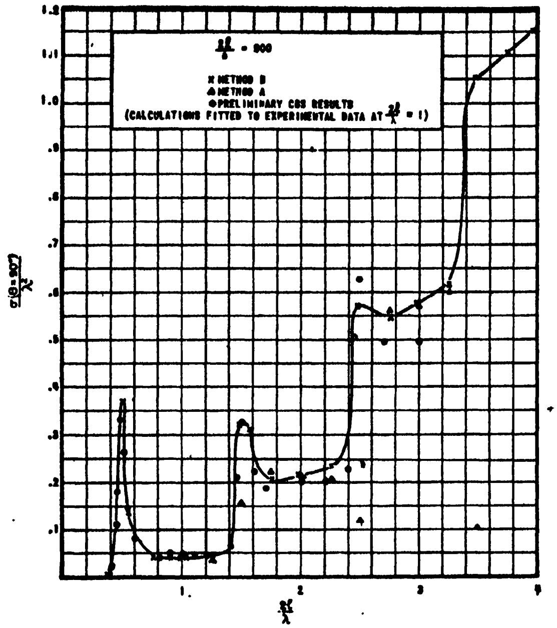


Fig. 4. Broadside response of wires.

(See Eq. (36) of Sec. III, where one will find a more complete discussion of expressions for the resonance maxima, including formulas (36′) and (36″) which are more accurate than (6) under certain conditions.) Here the cotangent applies to the resonances 1, 3, 5, etc., and the tangent to 2, 4, 6⋯.

The positions of the maxima furnished by method (B) are in quite good accord with experiment. For $2l/a=900$, it predicts that the successive maxima should come at $4l/\lambda=0.95$, 1.94, $2.93\underline{5}$, 3.93, $4.92\underline{5}$ whereas the observed values are $4l/\lambda=0.96$, 1.88, 2.94, 3.90, 4.92.

(3) *Influence of Forced Terms at Resonance-Broadside Response:* If we consider only the averaged cross section, the amplitude at resonance is determined almost entirely by the resonant term ($C_2$ or $C_4$ in Eq. (2)) and but little harm is done in omitting the "forced" members $C_1$ and $C_3$ in calculating the amplitudes of the first few peaks ($m=1$, 2, 3). When however, we do not average over all orientations of the target, but instead consider only the broadside response (at $\theta=90°$), this is no longer true. The theory is particularly simple for the broadside case, as the phase of the incident electric vector is the same for all parts of the wire. The curve furnished by method (B) for the cross section as a function of $2l/\lambda$ for broadside setting of the wire is shown in Fig. 4.[8] It is in quite good agreement with the experimental data indicated by circles. On the other hand, method (A) gives results not in accord with experiment at the odd resonances. (The even resonances have no particularly large broadside scattering.) The reason why the forced vibrations cannot be neglected, as in method (A), for broadside response at resonance, especially at the higher maxima, is the following:

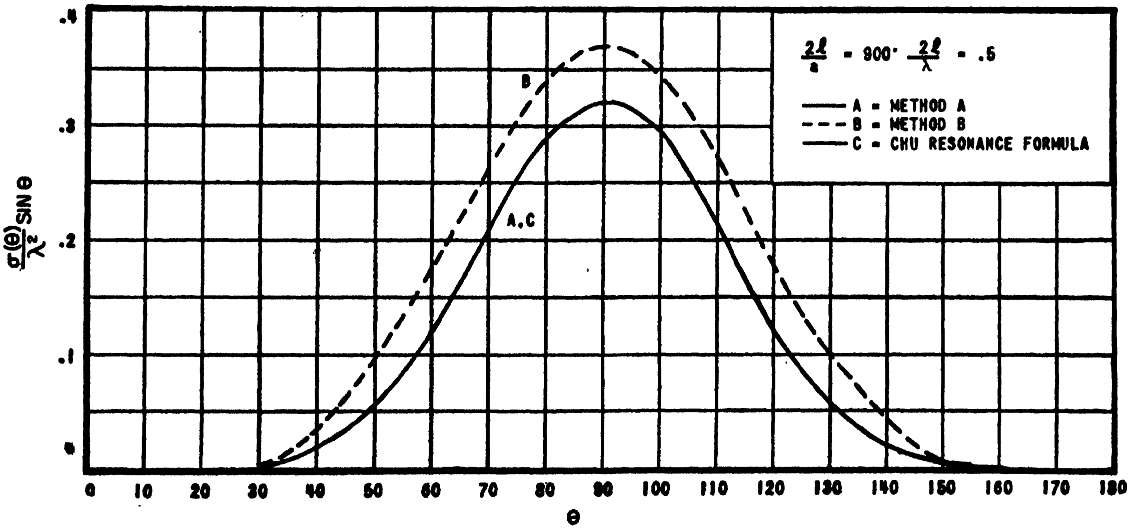


Fig. 5. Angular distribution of response of wire.

---

[8] The quantity plotted in Figs. 4–8 is $\sigma(\theta)$ which is connected with $\sigma$ by the relation $\sigma=(8/3)\cos^4\varphi\,\sigma(\theta)$.

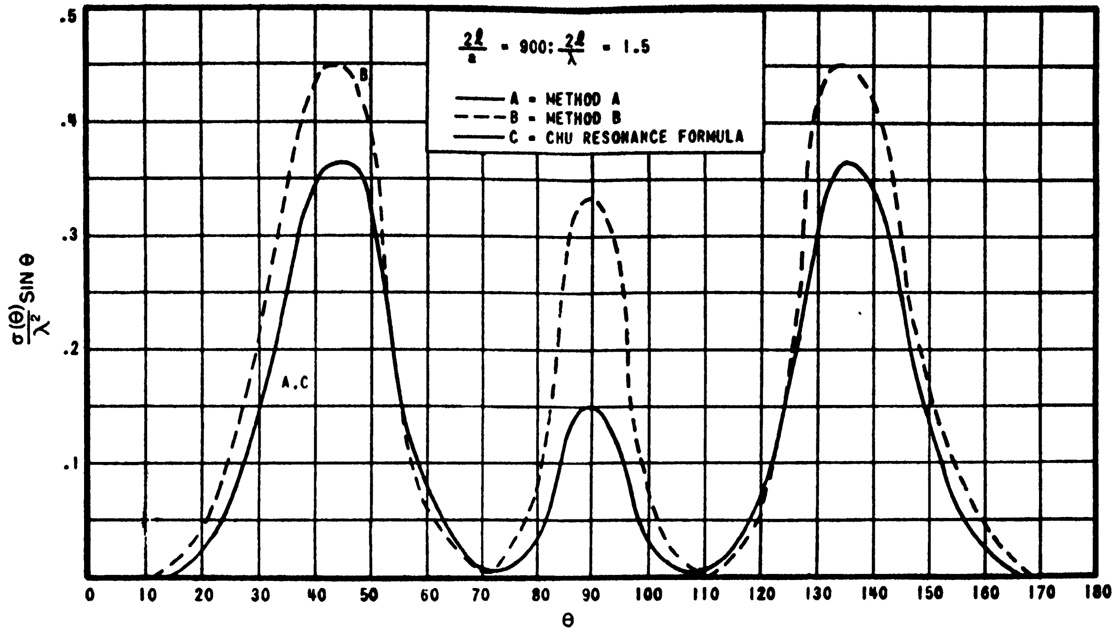


FIG. 6. Angular distribution of response of wire.

When we average over all orientations, the important term at resonance is usually that corresponding to resonant rather than forced oscillation (i.e., of structure $\sin\beta z$ or $\cos\beta z$ rather than $\sin qz$ or $\cos qz$ in Eq. (2)). In the broadside case, however, we are interested in scattering perpendicular to the wire (the equatorial zone). The forced oscillations $C_1$, $C_3$ in Eq. (2) give a particularly strong perpendicular scattering, whereas except for the first resonance or two, the major lobes due to the resonant vibrations $C_2$, $C_4$ are nearer the poles than the equator. For this reason the forced radiation is abnormally important in the broadside case, and at the higher resonances actually exceeds the resonant, despite the fact that the latter is associated with a far larger current amplitude. As an example we cite some numerical results obtained by method (B) for a wire having $2l/a=900$. Here at the first, third, fifth, seventh resonances, the contribution due to the resonant oscillations represents respectively 95, 54, 24, 11 percent of the total broadside cross section, the remainder being caused by the forced oscillations. It is because method (A) artificially suppresses the forced vibrations at resonance that it actually gives minima rather than maxima in the broadside response curves at resonance.

(4) *Height of the Maxima in the Averaged Cross Section:* The height of the resonance peaks as given by method (B) is higher than that given by method (A) or by the Chu formula. For instance, if $2l/a=900$, the maxima corresponding to $m=1$, 3, 5, 10, 20 are, respectively, 20, 32, 42, 66, 112 percent higher than given by Chu's expression. This difference, which would disappear for infinitely thin wires, is ascribable to two causes. One is the neglect of forced oscillations by (A), or by the Chu expression (our Eq. (4)), at resonance maxima. We have already discussed in paragraph (3) how this affects the results at broadside. For the first two or three maxima, the peaks of the averaged cross section are not affected appreciably by the forced oscillations. However, if, for example, $2l/a=900$, inclusion of the forced oscillations in method (B) makes the first, third, and tenth maxima ($m=1$, 3, 10), respectively, 6, 11, 29 percent greater than if the forced members are neglected (i.e., if $C_1$ and $C_3$ are taken equal to zero in Eq. (2)). The other effect which makes the maxima given by (B) higher than those given by (A) or by Chu is the fact that (B) takes into account the displacement of resonance to a slightly longer wave-length than that furnished by the ordinary expression $l/\lambda=m/4$. The theory of Part III shows that the

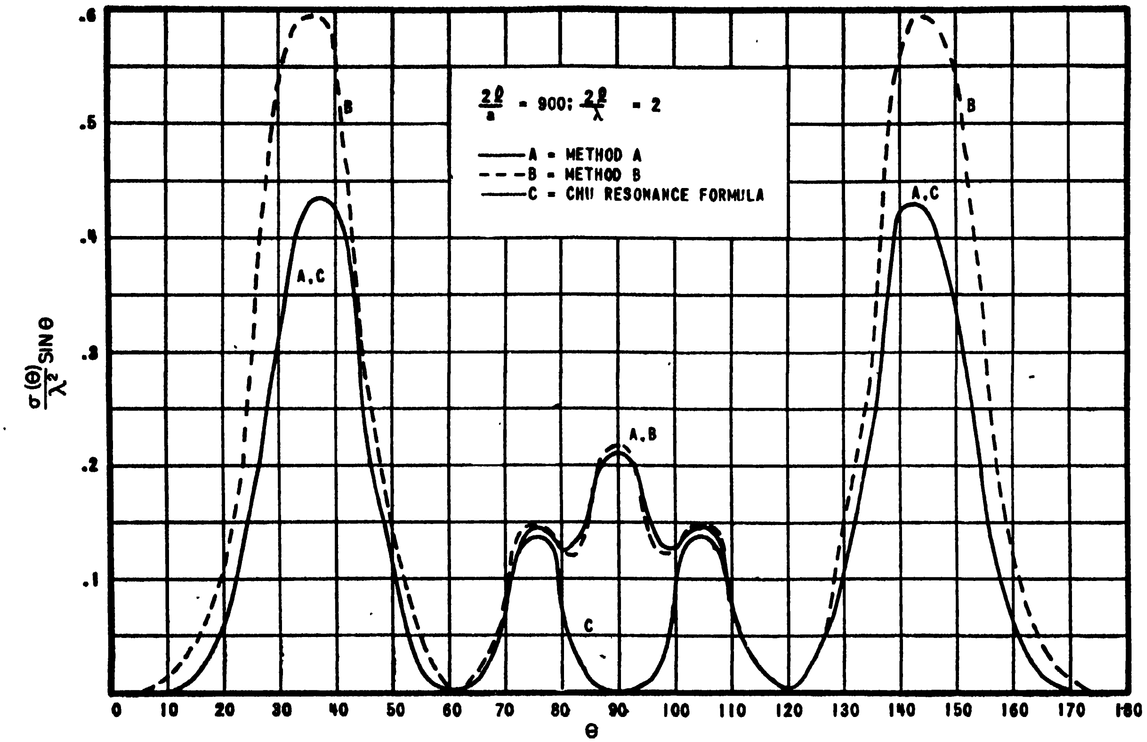


FIG. 7. Angular distribution of response of wire.

shift in the resonance decreases the damping and so increases the amplitude.[9] It is particularly noteworthy that this displacement effect makes the peak for $m=1$ higher than it would be otherwise, since this particular peak is often used for calibration. (Equation (4) shows that according to Chu the cross section at the first peak is $0.153\lambda^2$, regardless of the value of $l/a$. With method (B), on the other hand, it depends on the radius of the wire, and has the values of $0.184\lambda^2$, $0.194\lambda^2$, $0.204\lambda^2$ for $2l/a=900$, 450, 225, respectively.)

(5) *Height of the Valleys:* Here methods (A) and (B) agree quite well. The cross section furnished by the Chu non-resonant formula (Eq. (5a)), is indicated by the dotted line in Fig. 3. We see that even the deepest parts of the valleys are considerably higher than the dotted line. This means that even as far removed as possible from resonance, the effects of the resonant terms are not completely negligible. Stated in a different way, a finite wire, even when it is non-resonant, is far from behaving like an infinite wire. It should be noted that the valley floors depend on the radius much more than do the peaks.

(6) *Angular Distribution:* It is interesting to plot $\sigma(\theta)$ for a fixed wave-length. This has been done for the first, third, and fourth resonance maxima in Figs. 5–7. The types of angular variation predicted by methods (A) and (B) are not much different and are in general agreement with experiment. The angular patterns given by method (A) for the first few resonances are almost identical with those obtained by the simple theory of Chu. For higher maxima such as the 2 peak, however, the corrections to the simple theory are important, as shown by Fig. 7, especially at 90°.

As typical of an angular response well removed from resonance, the angular pattern for $2l/\lambda=1.25$ is given in Fig. 8, for which methods (A) and (B) give the same distribution.

(7) *Comparison with Experiment:* Unfortu-

---

[9] Mathematically, this is expressed in the fact that the "resonance denominators" $\Psi^2+\Xi^2$ appearing in the expressions for $G$ and $H$ in Eq. (35) are smaller in numerical magnitude for $\beta l<(m+\frac{1}{2})\pi$ and $\beta l<m\pi$, respectively, than for $\beta l=(m+\frac{1}{2})\pi$ and $\beta l=m\pi$.

nately, one will not find in the present paper elaborate curves comparing theory with experiment. This is because the existing data are, in our opinion, not yet satisfactory. In the absence of well-established experimental material, obviously more reliance must be placed on theory than would otherwise be the case, and this is one reason why we have developed the theory in detail in the two forms, (A) and (B); it is perhaps reassuring to have two independent theoretical approaches. Two results of the theory, for which significant comparison with experiment can be made at present have already been discussed, *viz.*, (a) shift in the positions of the resonance maxima from the "zero-order" positions $l=\frac{1}{4}m\lambda$; and (b) broadside response. Experiments on either of these features are easier than on the averaged cross section. Obviously (a) does not involve any intensity measurements, while (b) does not involve a variety of angular settings such as are needed in finding the averaged cross section. A more important feature of measurements of the broadside response is that they avoid polarization effects in the suspensions used to support the wires during laboratory measurements. The dielectric properties of the suspension may alter the charge distribution and hence the current in the wire. However, in the broadside case the current is a symmetrical function of the distance from the center of the wire, so that at the center we have $\partial I/\partial z=0$. Furthermore, the equation of continuity gives $\partial I/\partial z=i\omega\rho$, and hence there will be no charge near the center of the wire. Thus the polarization difficulties will not occur if the wire is supported at its center.

*Limitations of the Theory:* It is essential that the reader should appreciate that the antenna problem is so difficult that even these somewhat involved calculations apply only under the assumption that the wire is long compared to its radius. The calculations involve an expansion in descending powers of $\Omega=2\log(2l/a)$, where $l$ is the half-length of the wire, and $a$ is the radius of the equivalent wire. The fact that the expansion is in terms of reciprocal powers of $\log 2l/a$, rather than of $2l/a$, itself shows that under no conditions can the accuracy be monumental. For values of $2l/a$ of the order $10^3$ it is doubtless quite good (say within a few percent) and for $2l/a\sim 100$ it is probably adequate, but for very "fat" wires of the order $2l/a=10$ the calculations can at most be relied on only for orders of magnitude. Even when $l/a$ is very large, the theory may involve considerable percentage of error in its expression for the amplitude of the resonant vibrations, if the wire is so very long as to contain

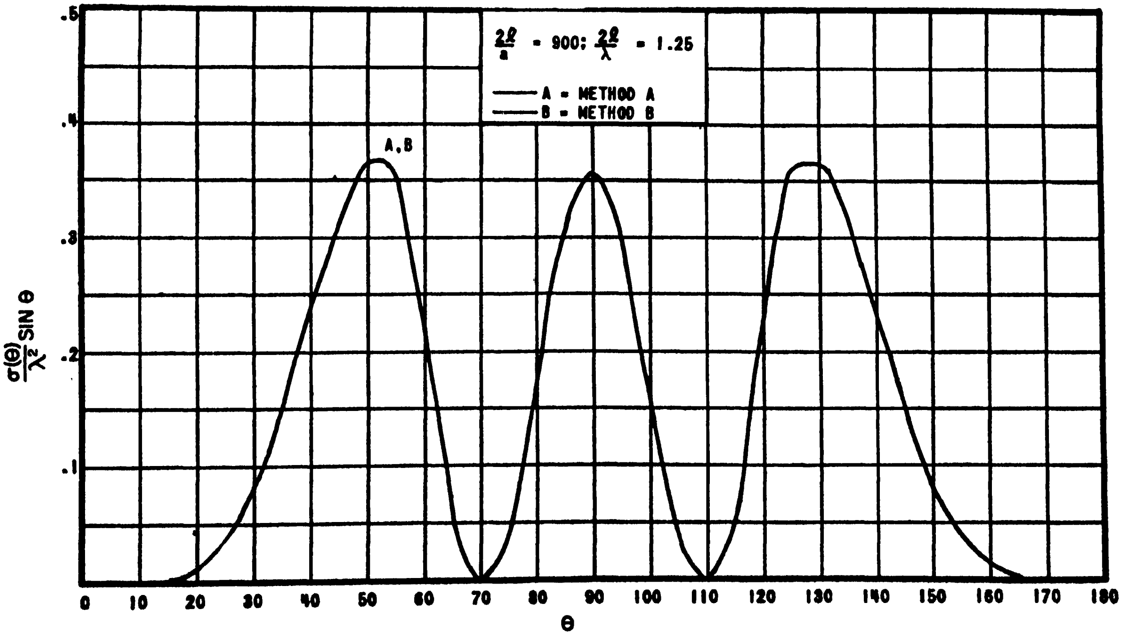


FIG. 8. Angular distribution of response of wire.

a large number of wave-lengths (see discussion in latter part of appendix). However, this difficulty is not serious, because for such exceedingly long wires the resonant vibrations are of minor importance compared to the forced ones.

## II. DETAILS OF METHOD A

The projection of the incident field on the antenna is $E_0 \sin\theta \cos\varphi e^{iqz}$. (We omit everywhere the time factor $e^{-i\omega t}$.) We choose our current in a form suggested by the solution of King and Harrison[3]:

$$I(z) = -\frac{i\omega E_0}{\beta^2 \sin\theta} \frac{\cos\varphi}{\Omega}\left[\frac{\cos qz \cos\beta l - \cos\beta z \cos ql}{\cos\beta l + (i/\Omega) f(l, q)} + i\frac{\sin qz \sin\beta l - \sin\beta z \sin ql}{\sin\beta l + (i/\Omega) g(l, q)}\right]; \tag{7}$$

$$\Omega = 2[\log(2/\gamma\beta a) + \mathrm{Ci}2\beta l]; \quad \gamma = 1.78; \quad \mathrm{Ci}x = -\int_x^\infty \frac{\cos t}{t} dt. \tag{8}$$

In Eq. (7) we shall refer to the first term in brackets as the even part and to the second as the odd part. To determine the functions $f$ and $g$ we use the procedure of calculating the work done on the antenna, and equate to the total energy flux radiated from the antenna; i.e., we set

$$\mathrm{Re}\int_{-l}^{+l} \mathbf{E} \cdot \mathbf{I}^* dz = \frac{c}{4\pi} \mathrm{Re} \int (\mathbf{E} \times \mathbf{H}^*) R^2 d\omega = \frac{c}{4\pi} \int |\mathbf{E}|^2 R^2 d\omega, \tag{9}$$

where $\mathbf{E}$, $\mathbf{H}$ are the fields produced by the induced currents in the antenna, $F^*$ is the conjugate complex of $F$, and Re denotes the real part. The first integral is taken over the antenna, while the second is taken over the surface of a large sphere of radius $R$.

On the antenna we have

$$\int_{-l}^{l} \mathbf{E} \cdot \mathbf{I}^* dz = -\frac{i\omega E_0{}^2 \cos^2\varphi}{\beta^2 \Omega}\left[\frac{\cos\beta l\{l + h(2q)\} - \cos ql\{h(q+\beta) + h(q-\beta)\}}{\cos\beta l - (i/\Omega) f(l, q)}\right.$$
$$\left. + \frac{\sin\beta l\{l - h(2q)\} - \sin ql\{h(q-\beta) - h(q+\beta)\}}{\sin\beta l - (i/\Omega) g(l, \beta)}\right] \tag{10}$$

with the abbreviation $h(x) = \sin xl/x$. Again we shall refer to the first term as the even part, and to the second as the odd part.

The field produced by the current $I$ at $\measuredangle\theta'$ and distance $R$ is

$$E = \frac{i\beta}{c} \frac{\sin\theta'}{R} \int_{-l}^{+l} I(z) e^{i\beta z \cos\theta'} dz. \tag{11}$$

The flux density is

$$\frac{c}{4\pi}|E|^2 = \frac{c}{4\pi} \frac{\beta^2}{c^2} \frac{\sin^2\theta'}{R^2} \int_{-l}^{+l} \int_{-l}^{+l} I(z) I^*(z') e^{i\beta(z-z') \cos\theta'} dz dz'. \tag{12}$$

The total flux is

$$\frac{c}{4\pi} \int_0^{2\pi} d\varphi' \int_0^\pi \sin\theta' d\theta' |\mathbf{E}|^2 R^2 = \frac{1}{\omega}\left\{ \int_{-l}^{+l}\int I^*(z') \frac{\sin\beta(z-z')}{z-z'}\left[\left(\beta^2 + \frac{d^2}{dz^2}\right) I(z)\right] dz dz'\right.$$
$$\left. - \int_{-l}^{+l} dz' I^*(z') \left( \frac{dI(z)}{dz} \frac{\sin\beta(z-z')}{z-z'} \Big|_{z=-l}^{z=+l} \right) \right\}. \tag{13}$$

We have written the current as the sum of an even and an odd function of $z$. As can be seen from Eq. (13), integrals involving products of even and odd parts of the current vanish; i.e., the total flux consists of two parts, one in which we choose for $I(z)$ only the even part of the current; the other in which we use only the odd part.

We shall satisfy Eq. (9) (which is actually simply a statement of the law of conservation of energy) separately for the odd and even parts. Using only the even part of $I(z)$ from Eq. (7) in Eq. (13), and performing the integrations, we obtain for this part of the total flux the value

$$\frac{1}{\cos^2\beta l+f^2/\Omega^2}\,\frac{1}{\omega}\,\frac{\omega^2 E_0^2\cos^2\varphi}{\beta^4\sin^2\theta\,\Omega^2}\left[\beta\gamma(\beta)\cos^2 ql+K\sigma(q)-L\gamma(q)-M\cos^2\beta l\right], \tag{14}$$

where

$$K=(\beta^2-q^2)\{l+h(2q)\}\cos^2\beta l+2\cos\beta l\cos ql(q\sin ql\cos\beta l-\beta\cos ql\sin\beta l),$$

$$L=(\beta^2-q^2)\cos^2\beta l(\cos 2ql/2q)+q\cos^2\beta l\cos 2ql+\tfrac{1}{2}\beta\sin 2ql\sin 2\beta l,$$

$$M=(\beta+q)\sin^2(\beta-q)l+(\beta-q)\sin^2(\beta+q)l,\quad \sigma(q)=\mathrm{Si}2(\beta+q)l+\mathrm{Si}2(\beta-q)l,$$

$$\gamma(q)=\mathrm{Cin}2(\beta+q)l-\mathrm{Cin}2(\beta-q)l,\quad \mathrm{Si}x=\int_0^x\frac{\sin t}{t}dt;\quad \mathrm{Cin}x=\int_0^x\frac{1-\cos t}{t}dt=\log x+0.577-\mathrm{Ci}x.$$

Equating (14) to the even part of Eq. (10), and solving for $f(l, q)$, we obtain

$$f=\sigma(y)\cos x+A^{-1}\{\gamma(x)\cos^2 xy-\gamma(y)[\tfrac{1}{2}y^{-1}(1+y^2)\cos^2 x\cos 2xy+\tfrac{1}{2}\sin 2x\sin 2xy]$$
$$-\cos^2 x[(1+y)\sin^2 x(1-y)+(1-y)\sin^2 x(1+y)]\}, \tag{15}$$

where

$$x=\beta l;\quad y=\cos\theta=q/\beta;\quad \sigma(y)=\mathrm{Si}2x(1+y)+\mathrm{Si}2x(1-y);\quad \gamma(y)=\mathrm{Cin}2x(1+y)-\mathrm{Cin}2x(1-y);$$
$$\gamma(x)=\mathrm{Cin}4x;\quad A=(1-y^2)(x+\tfrac{1}{2}y^{-1}\sin 2xy)\cos x-\cos xy\{(1+y)\sin x(1-y)+(1-y)\sin x(1+y)\}.$$

By a similar procedure, using the odd part of the current, we obtain

$$g=\sigma(y)\sin x+B^{-1}\{\gamma(x)\sin^2 xy+\gamma(y)[\tfrac{1}{2}y^{-1}(1+y^2)\sin^2 x\cos 2xy-\tfrac{1}{2}\sin 2x\sin 2xy]$$
$$-\sin^2 x[(1+y)\sin^2 x(1-y)+(1-y)\sin^2 x(1+y)]\}, \tag{16}$$

where

$$B=(1-y^2)(x-\tfrac{1}{2}y^{-1}\sin 2xy)\sin x-\sin xy\{(1+y)\sin x(1-y)-(1-y)\sin x(1+y)\}.$$

Having determined the functions $f$ and $g$, we now calculate the scattering cross section in the backward direction $\theta$. The scattering cross section is defined as $\sigma=4\pi R^2|E|^2/|E_0|^2$ when $E$ is the scattered field in direction $\theta$, as given by Eq. (11). After performing the integration, we can write $E$ in terms of $x$ and $y$ as

$$E=\frac{E_0\cos\varphi}{\beta\Omega(1-y^2)}\left[\frac{A}{\cos x+if/\Omega}-\frac{B}{\sin x+ig/\Omega}\right], \tag{17}$$

$$|E|^2=\frac{E_0^2\cos^2\varphi}{\beta^2\Omega^2(1-y^2)^2}\left[\frac{A^2}{\cos^2 x+f^2/\Omega^2}+\frac{B^2}{\sin^2 x+g^2/\Omega^2}-\frac{2AB(\sin x\cos x+fg/\Omega^2)}{(\sin x\cos x+fg/\Omega^2)^2+\Omega^{-2}(g\cos x-f\sin x)^2}\right]. \tag{18}$$

We assume that both transmitter and receiver are polarized parallel to each other. This gives an extra factor $\cos^2\varphi$. The average over all polarizations give $\langle\cos^4\varphi\rangle_{\mathrm{Av}}=\tfrac{3}{8}$ (if the polarizations of receiver and transmitter were crossed, we would be concerned with $\langle\sin^2\varphi\cos^2\varphi\rangle_{\mathrm{Av}}=\tfrac{1}{8}$, making the cross section one-third as large). Thus we have

$$\frac{\sigma(\theta)}{\lambda^2}=\frac{3}{8\pi\Omega^2}\frac{1}{(1-y^2)^2}\left[\frac{A^2}{\cos^2 x+f^2/\Omega^2}+\frac{B^2}{\sin^2 x+g^2/\Omega^2}-\frac{2AB(\sin x\cos x+fg/\Omega^2)}{(\sin x\cos x+fg/\Omega^2)^2+\Omega^{-2}(g\cos x-f\sin x)^2}\right]. \tag{19}$$

TABLE I.

| $2l/\lambda$ | 1/4 | 1/2 | 3/4 | 1 | 5/4 | 3/2 | 7/4 | 2 | 15/4 |
|---|---|---|---|---|---|---|---|---|---|
| $\bar{\sigma}/\lambda^2$ | $6\times10^{-5}$ | .151 | .023 | .163 | .022 | .176 | .066 | .219 | .127 |

Assuming that the wires are randomly oriented in space, the average cross section, $\bar{\sigma}$, is

$$\bar{\sigma}=\int_0^{\pi/2} d\theta\,\sin\theta\,\sigma(\theta), \tag{20}$$

where $\sigma(\theta)$ is given by Eq. (19). The evaluation of $\bar{\sigma}/\lambda^2$ requires numerical integration of Eq. (20). For $2l/a=900$, the results are given in Table I.

The parameter $\Omega$ increases as the radius of the antenna decreases. The terms of Eq. (19) are of order $1/\Omega^2$, except when $\cos x$ or $\sin x$ become small; i.e., the cross section will be small ($\sim 1/\Omega^2$) except for values of $x$ nearly $\pi/2$, $3\pi/2$, etc. (odd resonances) and $\pi$, $2\pi$, etc. (even resonances). Figures 5 to 8 show plots of $\sigma(\theta)/\lambda^2$ as a function of $\theta$ as computed from Eq. (19).

The broadside cross section is of interest for comparison with experiment and with method (B). For this case $\theta=90°$, $y=0$. If we let $y$ approach zero, then $A\rightarrow 2x\cos x-\sin x$; $B\rightarrow 0$; $f\rightarrow 2\cos x\,\mathrm{Si}2x+\frac{1}{2}[\gamma(x)-\sin^2 2x](x\cos x-\sin x)^{-1}$; $g$ remains finite, so that the odd term gives no contribution broadside. This is easily seen physically. For an odd distribution of current, contributions to the field broadside will cancel in pairs.

Substituting in Eq. (19) we obtain

$$\frac{\sigma(\theta=90°)}{\lambda^2}=\frac{3}{2\pi}\,\frac{(x\cos x-\sin x)^2}{\Omega^2\cos^2 x+\{2\cos x\,\mathrm{Si}2x+\frac{1}{2}[\gamma(x)-\sin^2 2x][x\cos x-\sin x]^{-1}\}^2}. \tag{21}$$

For $\theta=90°$, $\sigma(\theta)/\lambda^2$ is plotted in Fig. 4, for $2l/a=900$. Equation (19) reduces, in limiting cases, to the results obtained by Chu. At odd resonances ($x=\pi/2$, $3\pi/2$, etc.) $\sin x=1$, $\cos x=0$. Then the contribution of the odd terms will be of order $1/\Omega^{-2}\times$ (even terms). If we neglect the odd terms then we find

$$\sigma(\theta)/\lambda^2=\tfrac{3}{8}\pi^{-1}[A/f(1-y^2)]^2,\quad\text{with}\quad f^2=\tfrac{1}{4}[\gamma(x)]^2;\quad A^2=4\cos^4 xy.$$

The corresponding averaged cross section is

$$\bar{\sigma}/\lambda^2=\tfrac{3}{2}\pi^{-1}\gamma(x)^{-2}[x(\mathrm{Si}4x-\tfrac{1}{2}\,\mathrm{Si}8x)+\tfrac{1}{2}\,\mathrm{Cin}4x-\tfrac{1}{8}\,\mathrm{Cin}8x]. \tag{22}$$

Using the asymptotic expansions $\mathrm{Si}z\sim\frac{1}{2}\pi-(\cos z/z)$; $\mathrm{Cin}z\sim\log\gamma z+\sin z/z$; we have

$$\frac{\bar{\sigma}^2}{\lambda^2}\simeq\frac{3}{16\pi}\,\frac{4\pi^2 l/\lambda-\frac{3}{2}+3\log(8\pi l\gamma/\lambda)-\log 2}{[\log(8\pi l\gamma/\lambda)]^2}, \tag{22a}$$

which is the expression obtained by Chu (Eq. (4)). A similar derivation for the even resonances yields the same final formulas (22) and (22a).

For long wires ($x\rightarrow\infty$), we have, provided we are not at resonance and so long as $y\neq 0$:

$$f\rightarrow\sigma(y)\cos x;\quad g\rightarrow\sigma(y)\sin x;\quad A\rightarrow x(1-y^2)\cos x;\quad B\rightarrow x(1-y^2)\sin x.$$

Substituting in Eq. (19) we find that the return cross section is zero (for $y\neq 0$); i.e., for long non-resonant wires the return signal is a sharp broadside lobe.

For large values of $x$, one finds from Eq. (21),

$$\frac{\sigma(\theta=90°)}{\lambda^2}\simeq\frac{3}{2\pi}\,\frac{x^2}{\Omega^2+\pi^2}=\frac{3\pi l^2}{2}\,\frac{1}{[\log(2/\gamma\beta a)]^2+\pi^2/4}, \tag{23}$$

which is identical with the expression obtained by Chu (cf. Eq. (5), with $\theta=90°$ and $\langle\cos^4\varphi\rangle_{Av}=\frac{3}{8}$).

## III. DETAILS OF METHOD B

The fundamental integral equation to be satisfied after removal of the time factor is that already given as our Eq. (1). The constants $A_1$ and $A_2$ are to be determined by the requirement that the current vanishes at the end of the wire.

We seek to satisfy Eq. (1) approximately by a solution of the form

$$I(z)=\alpha e^{iqz}+\gamma_1\cos\beta z+i\gamma_2\sin\beta z. \tag{24}$$

The coefficient $\alpha$ is determined so as to cancel the part of the right side of Eq. (1) which is proportional to $e^{iqz}$. As in the paper of Gray, we write

$$\int_{-l}^{+l} r^{-1}I(\xi)e^{-i\beta r}d\xi=I(z)\int r^{-1}\cos\beta r\,d\xi+\int r^{-1}[I(\xi)-I(z)]\cos\beta r\,d\xi-i\int r^{-1}I(\xi)\sin\beta r\,d\xi. \tag{25}$$

The limits of integration are to be understood throughout to be $\pm l$. In $\int r^{-1}\cos\beta r d\xi$ we must take into account the finite radius of the wire (otherwise the integral diverges), and following the usual procedure of antenna theory take the origin of $r$ as at the center of the wire. In the other members, on the other hand, the integrals converge at $r=0$ so that the wire here may be regarded as of negligible thickness. For a more thorough discussion of this point see Miss Gray's paper;[5] it will be assumed that the reader has some familiarity with the procedure used there.

With these approximations we have

$$\int_{-l}^{+l}\frac{\cos\beta r}{r}d\xi=Z(z),\quad\text{with}\quad Z(z)=\log\frac{[(l+z)^2+a^2]^{\frac{1}{2}}+(l+z)}{[(l-z)^2+a^2]^{\frac{1}{2}}-(l-z)}-\mathrm{Cin}\beta(l+z)-\mathrm{Cin}\beta(l-z), \tag{26}$$

$$\begin{aligned}\int r^{-1}e^{iq\xi}e^{-i\beta r}d\xi=\tfrac{1}{2}e^{iqz}\big[&2Z(z)+2\,\mathrm{Cin}\beta(l-z)+2\,\mathrm{Cin}\beta(l+z)-\mathrm{Cin}(\beta+q)(l-z)-\mathrm{Cin}(\beta-q)(l+z)\\&-\mathrm{Cin}(\beta-q)(l-z)-\mathrm{Cin}(\beta+q)(l+z)-i\,\mathrm{Si}(\beta+q)(l-z)-i\,\mathrm{Si}(\beta-q)(l+z)\\&-i\,\mathrm{Si}(\beta-q)(l-z)-i\,\mathrm{Si}(\beta+q)(l+z)\big]+\tfrac{1}{2}e^{iqz}\big[\mathrm{Cin}(\beta+q)(l-z)\\&+\mathrm{Cin}(\beta-q)(l+z)-\mathrm{Cin}(\beta+q)(l+z)-\mathrm{Cin}(\beta-q)(l-z)\\&+i\,\{\mathrm{Si}(\beta+q)(l-z)+\mathrm{Si}(\beta-q)(l+z)-\mathrm{Si}(\beta+q)(l+z)-\mathrm{Si}(\beta-q)(l-z)\}\big].\end{aligned} \tag{27}$$

The coefficients of $e^{iqz}$ in Eq. (27) are slowly varying functions of $z$, and we shall replace them by their mean values. Actually the deviations from the mean values are considerable near the very ends of the wire. Consequently, it may appear to the reader that a very sizable error is committed by using mean values in the current formulas when we apply the boundary conditions at the end of the wire. This point will be discussed more fully in the appendix, where it will be shown that the objection is not a serious one. The second of the two factors enclosed in $\big[\;\;\big]$ in Eq. (27) is an odd function of $z$, so that its mean value is zero. Furthermore

$$\begin{aligned}\langle Z(z)\rangle_{\mathrm{Av}}&=\tfrac{1}{2}l\int Z(z)dz=2[\log(2l/a)+\log 2-\mathrm{Cin}2\beta l-(\sin 2\beta l/2\beta l)]\sim 2[\log(\lambda/\pi a)-0.577],\\\langle\mathrm{Cin}k(l+z)\rangle_{\mathrm{Av}}&=\mathrm{Cin}2kl+(\sin 2kl/2kl)-1\sim\log 2kl+0.577-1,\\\langle\mathrm{Si}k(l+z)\rangle_{\mathrm{Av}}&=\mathrm{Si}2kl+(1/2kl)[\cos 2kl-1]\sim\tfrac{1}{2}\pi.\end{aligned} \tag{28}$$

The approximations following the wavy lines are contingent on $kl$ being somewhat large compared to unity, and may be considered adequate for our purposes except when the wire is less than a half wave long. Even with wires a half wave-length or so long, the error involved in using the asymptotic formulas is by no means negligible; if, for instance $(2l/a)=900$, the cross section at the first resonance is raised six percent when the asymptotic expansions are not made. Since the formulas based on these asymptotic expansions are much easier to compute with than are those involving the Si and Cin functions, and since we are usually interested in wires at least half a wave long, we shall carry through the calculations on the assumption that the use of asymptotic approximations is legitimate. However, we shall later give some of the more complicated final expressions which are obtained when these approximations are not made. The expression (27) now becomes $Ke^{iqz}$, where

$$K=2\{\log(\lambda/\pi a)-0.577\}+2\log(1/\sin\theta)-i\pi.$$

The appropriate value of the constant $\alpha$ in Eq. (24) is thus

$$\alpha=i\omega\cos\varphi E_0/K\beta^2\sin\theta. \tag{29}$$

We now proceed to determine the relation between $\gamma_1$, $\gamma_2$ and the constants $A_1$, $A_2$ of Eq. (1). Using the same type of approximation as in Eq. (27) we obtain

$$\int r^{-1}\cos\beta\xi e^{-i\beta r}d\xi=L\cos\beta z;\quad \int r^{-1}\sin\beta\xi e^{-i\beta r}d\xi=L\sin\beta z,$$

where

$$L=2[\log(\lambda/\pi a)-0.577]+\log 2\beta l+0.577-\log 2-i\pi/2-1.$$

Thus

$$\gamma_1=A_1/L;\quad i\gamma_2=A_2/L. \tag{30}$$

We must now determine the constants $A_1$ and $A_2$ so as to satisfy the boundary condition $I(\pm l)=0$. Here it is not sufficient merely to set the expression (24) equal to zero at $z=\pm l$, which would give $\gamma_1=-\alpha\cos ql/\cos\beta l$; $\gamma_2=-\alpha\sin ql/\sin\beta l$, in other words infinite current at resonance. Instead we can obtain a finite current at resonance and proper damping and radiation resistance only by demanding that the second, rather than first approximation to the current, vanish at the boundary. This second approximation is obtained by iterating once the first approximation to the current, or in other words using the first approximation to $I(\xi)$ and $I(z)$ in the second and third, but not the first term in the right side of (25). Clearly the first term is much more sensitive than the other two to the value of $I(z)$, as it involves an integral which would be infinite if the wire radius, $a$, were equal to zero, whereas the other integrals always remain finite even with $a=0$. The boundary conditions for determining $A_1$ and $A_2$ can thus be taken to be

$$\int r^{-1}[I(\xi)-I(\pm l)]\cos\beta r d\xi-i\int r^{-1}I(\xi)\sin\beta r d\xi=(i\omega\cos\varphi/\beta^2\sin\theta)E_0e^{\pm iql}$$
$$+A_1\cos(\pm\beta l)+A_2\sin(\pm\beta l); \tag{31}$$

as is seen by substituting (25) in (1) with $z=\pm l$, and with $I(\pm l)=0$ in the first right-hand member of (25). The approximation (24) may be used for $I(\xi)$ and $I(l)$ in evaluating the left side of (31), and $r$ has the value $|l\pm\xi|$.

The satisfaction of the boundary conditions at the two ends of the wire is equivalent to satisfying the condition at one end separately for the problem of an "even" impressed field $E_0\cos qz$ and an "odd" one $iE_0\sin qz$, rather than lumping both in the combination $e^{iqz}$. Instead of regarding the boundary conditions as determining $A_1$ and $A_2$, we can equally well regard them as specifying $\gamma_1$, $\gamma_2$ inasmuch as $\gamma_1$, $\gamma_2$ and $A_1$, $A_2$ are related by Eq. (30); it is more convenient to solve for $\gamma_1$, $\gamma_2$ than $A_1$, $A_2$, since the former enter directly into the expression (24) for the current distribution. When we utilize (29), (30), and (27) (approximated in the usual way for large $l$) we thus find that (24) yields the following expressions for determining $\gamma_1$, $\gamma_2$:

$$(i\omega\cos\varphi E_0/K\beta^2\sin\theta)\{[\log(1/\sin\theta)-\tfrac{1}{2}i\pi]\cos ql-\tfrac{1}{2}i\log[(1+\cos\theta)/(1-\cos\theta)]\sin ql\}$$
$$+\gamma_1\{\cos\beta l(\tfrac{1}{2}\log 2\beta l+0.288-\tfrac{1}{2}\log 2-\tfrac{1}{4}i\pi)-\tfrac{1}{2}i(\log 4\beta l+0.577)\sin\beta l+\tfrac{1}{4}\pi\sin\beta l\}$$
$$=(i\omega\cos\varphi E_0/\beta^2\sin\theta)\cos ql+\gamma_1L\cos\beta l, \tag{32}$$

$$(i\omega \cos\varphi E_0/K\beta^2 \sin\theta)\{[\log(1/\sin\theta)-i\pi/2]\sin ql+\tfrac{1}{2}i\log[(1+\cos\theta)/(1-\cos\theta)]\cos ql\}$$
$$+\gamma_2\{\sin\beta l(\tfrac{1}{2}\log 2\beta l+0.288-\tfrac{1}{2}\log 2-\tfrac{1}{4}i\pi)+\tfrac{1}{2}i(\log 4\beta l+0.577)\cos\beta l-\tfrac{1}{4}\pi\cos\beta l\}$$
$$=(i\omega\cos\varphi E_0/\beta^2\sin\theta)\sin ql+\gamma_2 L\sin\beta l. \quad (33)$$

We may, without too much error, replace $\log(1/\sin\theta)$ and $\log(1+\cos\theta)/(1-\cos\theta)$ by zero in these expressions. The justification for this is that the bulk of the scattering comes from $\theta>30°$, so that the omitted terms are unimportant, and their inclusion would excessively complicate the angular dependence and preclude our obtaining analytical expressions for the cross section averaged over angle. We shall also similarly neglect the term $\log(1/\sin\theta)$ in the quantity $K$ of Eq. (29). Of course, these minor angular terms could be included, but they yield nothing worth while. The expression for the current may now, in virtue of (24), (29), (32), (33) be written as

$$I(z)=(i\omega\cos\varphi E_0/\beta^2\sin\theta)[(F'+iF'')e^{iqz}+2(G'+iG'')\cos\beta z\cos ql+2i(H'+iH'')\sin\beta z\sin ql], \quad (34)$$

$$F'=\Omega'/(\Omega'^2+\pi^2),\quad F''=\pi/(\Omega'^2+\pi^2),\quad \Omega'=2\log(\lambda/\pi a)-1.154, \quad (34a)$$

$$\left.\begin{aligned}2G'&=\Psi(\beta l)[\Psi(\beta l)^2+\Xi(\beta l)^2]^{-1}-\pi\Omega'^{-1}G'',\quad 2G''=\Xi(\beta l)[\Psi(\beta l)^2+\Xi(\beta l)^2]^{-1},\\ 2H'&=\Psi(\beta l-\tfrac{1}{2}\pi)[\Psi(\beta l-\tfrac{1}{2}\pi)^2+\Xi(\beta l-\tfrac{1}{2}\pi)^2]^{-1}-\pi\Omega'^{-1}H'',\end{aligned}\right\} \quad (35)$$

$$2H''=\Xi(\beta l-\tfrac{1}{2}\pi)[\Psi(\beta l-\tfrac{1}{2}\pi)^2+\Xi(\beta l-\tfrac{1}{2}\pi)^2]^{-1},$$
$$\Psi(x)=-(\Omega'-\Delta)\cos x+\tfrac{1}{4}\pi\sin x,\quad \Xi(x)=\tfrac{1}{2}(\log 4\beta l+0.577)\sin x-\tfrac{1}{4}\pi\cos x,\quad \Delta=-\tfrac{1}{2}\log\beta l+0.712. \quad (35a)$$

The terms $-\pi G''/\Omega'$, $-\pi H''/\Omega'$ occurring at the very ends of our formulas for $G'$ and $H'$ are important only very near resonance, where $G''$ and $H''$ are abnormally large. In writing down these terms, we have neglected $\Delta$, etc., in comparison with $\Omega'$ so that their denominators are simply $\Omega'$. This approximation is allowable as the terms $-\pi G''/\Omega'$, $-\pi H''/\Omega'$ are never of dominant importance.

Resonance occurs when there is a peak in the cross section plotted against wave-length. If the wire is sufficiently thin, so that $2l/a$ is of the order $10^3$, the only important terms which vary acutely with wave-length are the $G$ terms at the resonances 1, 3, 5 ⋯ and the $H$ ones at 2, 4, 6. The problem of locating the resonance is that of finding the position of the maximum of $G'^2+G''^2$ or $H'^2+H''^2$, and resonance can be regarded as occurring when the denominator of the fractions in Eq. (35) is as small as possible. If the wire is not too long, or in other words if $\log\beta l$ can be regarded as very small compared to $2\log(2l/a)$, the resonance is located nearly at the point where the first of the two squared terms composing the denominator is equal to zero. Then the forced and resonant vibrations are 90° out of phase. With this approximation, resonance occurs when the factor $(\Omega'-\Delta)\cos\beta l-\frac{1}{4}\pi\sin\beta l$, or $(\Omega'-\Delta)\sin\beta l+\frac{1}{4}\pi\cos\beta l$ vanishes. The positions of the odd and even resonances are thus not given by $\cos\beta l=0$ or $\sin\beta l=0$, as commonly assumed, but rather by

$$\cot\beta l=\pi/4(\Omega'-\Delta)\quad\text{or}\quad -\tan\beta l=\pi/4(\Omega'-\Delta). \quad (36)$$

This relation is the same as Eq. (6) of Part I. If asymptotic expansions are not used, Eq. (36) is replaced by

$$\begin{pmatrix}\cot\beta l\\ -\tan\beta l\end{pmatrix}=\frac{\tfrac{1}{2}\,\mathrm{Si}4\beta l}{2[\log(2l/a)+\log 2-1]-\tfrac{1}{2}\,\mathrm{Cin}4\beta l-\mathrm{Cin}2\beta l-(\sin 4\beta l/4\beta l)+1}. \quad (36')$$

Substitution of (36′) for (36) alters the resonance frequency only negligibly except at the first maximum, and even here the change is slight, (e.g., 0.3 percent in $\beta l$ if $2l/a=900$). A more important refinement is the allowance for the fact that the minima of the denominators of $G$ or $H$ are not located exactly where their first terms vanish. Correction for this fact leads to the formula

$$\begin{pmatrix}\cot\beta l\\ -\tan\beta l\end{pmatrix}=\frac{\pi(1+\epsilon)}{4(\Omega'-\Delta)}\quad\text{with}\quad \epsilon=\frac{[\log(8\pi l/\lambda)+0.577]^2+2[\log(8\pi l/\lambda)+0.577][\Omega'-\Delta]}{[2(\Omega'-\Delta)]^2}. \quad (36'')$$

The difference between (36) and (36″) is particularly noticeable if the wire is long enough to contain several wave-lengths. For instance if $2l/a=900$, use of (36″) rather than (36) increases the shift of the first peak (measured from the Chu value) by only 10 percent, but the 10th peak by 28 percent. The numerical figures quoted in Sec. I are based on (36″) rather than (36), and in the first peak, avoidance of asymptotic expansions in the leading term is avoided by inserting the correction factor $1+\epsilon$ in (36′) rather than (36). It must be cautioned that if $2l/a$ is of the order $10^2$, so that the convergence of our method becomes slow due to the inadequate size of $\Omega$, none of the formulas (36), (36′), (36″) should be used for locating the maximum, as then the influence of the cross terms $FG$ or $FH$ becomes important and the resonance must be located by plotting the final expression (38) for the averaged cross section as a function of wave-length rather than merely maximizing $G'^2+G''^2$ or $H'^2+H''^2$. Thus if $2l/a=80$ it turns out that the shift given by (36″) is about 50 percent larger than what is obtained by locating the peak in the complete expression for the cross section.

The reflected field can be determined from $I(z)$ as given by Eq. (34) by following the same procedure as in Eqs. (11) and (12). We thus obtain

$$\begin{aligned}\sigma&=\frac{\cos^2\varphi(4\pi R^2)|E|^2}{|E_0|^2}\\&=4\pi\cos^4\varphi\left|\int_{-l}^{+l}dze^{iqz}[(F'+iF'')e^{iqz}+2(G'+iG'')\cos\beta z\cos ql+2i(H'+iH'')\sin\beta z\sin ql]\right|^2\\&=(4\lambda^2/\pi)\cos^4\varphi\{a_1^2(F'^2+F''^2)+(a_2+a_3)^2\cos^2 xy(G'^2+G''^2)+(a_2-a_3)^2\sin^2 xy(H'^2+H''^2)\\&\qquad+2(a_2^2-a_3^2)\sin xy\cos xy(G'H'+G''H'')+2a_1(a_2+a_3)\cos xy(F'G'+F''G'')\\&\qquad\qquad+2a_1(a_2-a_3)\sin xy(F'H'+F''H'')\},\end{aligned}\tag{37}$$

where

$$x=2\pi l/\lambda,\quad y=\cos\theta,\quad a_1=\tfrac{1}{2}y^{-1}\sin 2xy,\quad a_2=(1+y)^{-1}\sin[x(1+y)],\quad a_3=(1-y)^{-1}\sin[x(1-y)].$$

$\sigma(\theta)$ can be obtained from (37) by replacing $\cos^4\varphi$ by its mean value $\frac{3}{8}$. (For the case of crossed polarization see the remarks in Sec. II following Eq. (18).) To find the value of $\bar{\sigma}$, Eq. (37) should in addition be integrated from $y=0$ to $y=1$. (It must not be integrated from $-1$ to $+1$ as in evaluating the Si functions we have assumed their arguments positive, which requires $y>0$.)

The integrals are evaluable in terms of $\mathrm{Si}(x)$, $\mathrm{Ci}(x)$ and elementary functions.[10] It is convenient to make asymptotic expansions of $\mathrm{Si}(x)$ and $\mathrm{Ci}(x)$ after the integrals have been calculated. In some of the terms, however, it turns out that the factor multiplying $\mathrm{Si}(x)$ or $\mathrm{Ci}(x)$ contains a first power of $x$ explicitly rather than involving $x$ only through trigonometric functions, and in such terms it is not sufficient to use the approximations $\mathrm{Si}(x)\sim\frac{1}{2}\pi$, $\mathrm{Ci}(x)=0$, but rather we must employ the somewhat more accurate asymptotic developments $\mathrm{Si}(x)\sim\frac{1}{2}\pi-\cos x/x$, $\mathrm{Ci}(x)\sim\sin x/x$ in order to obtain answers correct to terms of the zeroth order in $x$. By proceding in this fashion it is found that the averaged cross section $\bar{\sigma}$ finally becomes

$$\begin{aligned}16\pi\bar{\sigma}/3\lambda^2=&(F'^2+F''^2)(2x\pi-1)+(G'^2+G''^2)[2x\pi-1+\log 4x+0.577+(4x\log 2+\pi)\sin 2x\\&\qquad+(\tfrac{1}{2}-2\{\log 4x+0.577\})\cos 2x-(\log 2)\cos 4x]\\&+(H'^2+H''^2)[2x\pi-1+\log 4x+0.577-(4x\log 2+\pi)\sin 2x\\&\qquad+(2\{\log 4x+0.577\}-\tfrac{1}{2})\cos 2x-(\log 2)\cos 4x]\\&+2(G'H'+G''H'')[4x(\log 2)\cos 2x-\tfrac{1}{2}\sin 2x]\\&+4(F'G'+F''G'')[(7\pi/4)\sin x-\tfrac{1}{2}(\log 4x+0.577)\cos x-\tfrac{1}{2}(\log 2)\cos 3x]\\&\qquad+4(F'H'+F''H'')[\tfrac{1}{2}\pi\cos x-\tfrac{1}{2}(\log 4x+0.577)\sin x+\tfrac{1}{2}(\log 2)\sin 3x].\end{aligned}\tag{38}$$

---

[10] For detailed evaluation of the integrals see report 411-103 of the Radio Research Laboratory, Harvard University (NDRC Div. 15). This report also contains more graphs of the angular cross sections than given in our Figs. 5–8.

*Short wires* in case the wire is not long enough compared to the wave-length to permit the use of asymptotic expansions, the preceding formulas must be modified as follows: Eq. (37) is still applicable, but the coefficients $F$, $G$, $H$, must be redefined. Equation (34a) is replaced by

$$F'=\Omega''/(\Omega''^2+\mathrm{X}^2),\quad F''=\mathrm{X}/(\Omega''^2+\mathrm{X}^2),\quad \mathrm{X}=2[\mathrm{Si}2\beta l+(2\beta l)^{-1}(\cos 2\beta l-1)],$$

$$\Omega''=2[\log(2l/a)+\log 2-\mathrm{Cin}2\beta l-(\sin 2\beta l/2\beta l)].$$

The formulas (35) for $G$ and $H$ are still valid if the coefficients $\pi\Omega'^{-1}$ in (35) are replaced by $2\Omega''^{-1}\mathrm{Si}2\beta l$, and if the definitions (35a) are modified to

$$\Psi(x)=-(\Omega''-\Delta'')\cos x+\tfrac{1}{2}\,\mathrm{Si}4\beta l\sin x,\quad \Xi(x)=\tfrac{1}{2}\,\mathrm{Cin}4\beta l\sin x-[\tfrac{1}{2}\mathrm{Si}4\beta l+(4\beta l)^{-1}(\cos 4\beta l-1)]\cos x,$$

$$\Omega''-\Delta''=2[\log(2l/a)+\log 2-1]-\tfrac{1}{2}\,\mathrm{Cin}4\beta l-\mathrm{Cin}2\beta l-(\sin 4\beta l/4\beta l)+1.$$

The explicit expression for the resulting averaged cross section is given in reference 10, and will not be reproduced here, as it is quite cumbersome.

For $2l/a=900$, the averaged cross section at the first resonance turns out to be six percent lower when asymptotic expansions are avoided. The lower values have been used in drawing the first peak in Figs. 2 and 3, but serious error would not have resulted had we used the asymptotic formulas. For the second and higher peaks, the results are not appreciably influenced by employing asymptotic expansions. On the other hand, when the frequency is lowered much below the first resonance, it is very essential to avoid using the asymptotic formulas.

When the frequency is quite small compared to the first resonant frequency, or in other words if the wire is cut much too short for resonance, it is very difficult to compute with the non-asymptotic formula that replaces (38). The cross section is expressed as a difference of terms which nearly cancel and which must hence be very accurately evaluated to obtain a significant answer.

However, for very short wires, simple formulas may be obtained by using for the coefficients the approximate values

$$F'=1/[2\log(2l/a)+2\log 2-1],\quad G'=-F'(1+x^2)/2,\quad H'=-F'/2x,$$

and developing the expression (34) for the current or the integrand of (37) as a Taylor's series in $x$ and retainining only the terms of lowest order. In this way one finds that before averaging the cross section is

$$\sigma=(\lambda^2x^6/9\pi)(\cos^4\varphi)(1-y^2)^2[\log(4l/a)-1]^{-2},\tag{39}$$

while the averaged cross section is

$$\bar{\sigma}=(\lambda^2x^6/45\pi)[\log(4l/a)-1]^{-2},\quad (x=2\pi l/\lambda).\tag{40}$$

Equations (39) and (40) can safely be used if the frequency is twenty percent or less of the first resonant frequency (i.e., $\lambda/l\geqq 20$). Equations (39) or (40) show that for wires very short compared to the wave-length, the cross section is proportional to the sixth power of the wire length, and inversely proportional to the fourth power of the wave-length. This behavior is reasonable, as in the low frequency limit the dimensions of the object and the wave-length also enter in the sixth and inverse fourth powers in the known formulas for other types of scatterers (e.g., spheres), as well as in the general Rayleigh theory of scattering by small particles.

We can, in fact derive (39) in a quite independent fashion by computing the Rayleigh scattering. It suffices to consider the case $\varphi=y=0$ where the wire axis $A$ is parallel to the incident vector $E_0$, as it is readily seen that for very long waves the cosine of the angle between $A$ and $E_0$ is involved to the fourth power, thus accounting for the factor $\cos^4\varphi(1-y^2)^2$ in (47). The broadside radiation of a dipole of moment $M$ has a Poynting vector $\frac{1}{4}\pi^{-1}c^{-3}R^{-2}(d^2M/dt^2)^2$. In the long wave-length limit we can safely take $d^2M/dt^2=-(2\pi\nu)^2a_0E_0$, where $a_0$ is the static polarizability of the wire. The cross section is then $a_0^2(2\pi/\lambda)^4 4\pi$. We now need only the value of $a_0$, and this we can take from the soluble problem[11] of a very prolate ellipsoid of revolution of semi-major axis and semi-minor axis $a$ in a static field parallel to the axis of symmetry. Since the wire is a conductor, we take the dielectric constant as $\infty$. The polarizability is then the volume of the ellipsoid divided by the demagnetizing factor $D$, and for the infinitely prolate case the value of $D$ is $(4\pi a^2/l^2)[\log(2l/a)-1]$. We thus obtain a formula identical with (39) except that $\log 2l/a$ occurs in place of $\log 4l/a$. The difference is of minor consequence since $2l/a$ is large, and is no doubt due either to the non-equivalence of an ellipsoid and cylinder, or to the fact that our replacement of $Z(z)$ by a constant $\langle Z\rangle_{\mathrm{Av}}$ as in (28) is of necessity an approximation.

The cross section falls off very rapidly as the frequency is lowered from the first resonance. Consequently in Fig. 2 or Fig. 3 the ordinate can be regarded as sensibly equal to zero when the abscissa is less than about half the value associated with the first resonance. Part of the drop of the curve as one approaches the origin is caused by the fact that $\bar{\sigma}/\lambda^2$ rather than $\bar{\sigma}$ is being plotted: $\lambda$ increases when the abscissa (proportional to frequency) is lowered. However, most of the decrease of $\bar{\sigma}/\lambda^2$ is caused by the $\bar{\sigma}$ rather than $1/\lambda^2$ factor. For instance, if $2l/a=900$, the value of $\bar{\sigma}/\lambda^2$ yielded by (40) at $x=0.3$ (frequency one-fifth that of the first resonance $x=1.5\sim\pi/2$) is 0.00000012, and the cross section $\bar{\sigma}$ is only one sixty-thousandth of its value at the first resonance. The values of $\bar{\sigma}/\lambda^2$ at frequencies forty, sixty, eighty, and one hundred percent of the first resonance ($x=0.6$, $x=0.9$, $x=1.2$, $x=1.50$) are 0.000043, 0.0014, 0.0071, 0.184, respectively, if $2l/a=900$.

## APPENDIX. SOME NOTES ON THE CONVERGENCE OF METHOD B

If we decompose the fundamental integral Eq. (1) into its forced and resonant parts, then the equation satisfied by either portion is mathematically of the type form

$$\int r^{-1} I_p(\xi) e^{-i\beta r} d\xi = D_p \cos kz, \tag{41}$$

where $I_p$ refers to only the forced or resonant part of the current ($p=f$, $r$, respectively) and where $k$ means $q=\beta\cos\theta$ for the forced and $\beta=2\pi/\lambda$ for the resonant case. For simplicity we consider only the portion of the solution which is even as regards reflection in the origin, as the convergence questions connected with the even and odd members are fundamentally similar. We seek to solve (41) by a trial function of the form $E_p \cos kx$. The left-hand side of (41) then turns out to be of the structure

$$E_p[f_p(z)\cos kz + g_p(z)\sin kz], \tag{42}$$

(Cf. Eq. (27)) where $f_p(z)$ is a slowly varying function of $z$, and $g_p(z)$ may be disregarded in the first approximation. We next replace $f_p$ by its mean value, and thus take

$$E_p = D_p/\bar{f}_p. \tag{43}$$

Against this procedure it may be objected by some readers that actually $f_p(z)$ does vary with $z$, and so should not be replaced by a constant in such a debonnaire fashion. In fact, at the ends of the wire $f(z)$ has approximately half its value at the center, which is also nearly its mean value, provided of course the wire length $l$ is large compared to its radius, as we assume throughout. Now for computation of the radiation from the wire, it seems fairly clear that the approximation $f_p=\bar{f}_p$ is not responsible for much error, since the ends of the wire contribute only a fraction of the total radiation. However, it may well seem at first sight that the use of (42) is a bad mistake in applying the boundary condition $I(\pm l)=0$, i.e., in satisfying Eq. (31), for it is at the ends of the wire that the difference between $f_p$ and $\bar{f}_p$ is largest. We shall now show that this objection is not as serious as it first seems. We may distinguish between two cases: (a) wires which contain at most a few wave-lengths, so that $\log(l/\lambda)$ is small compared with $\log(\lambda/a)$, and (b) wires which are so extended as to contain an enormous number of wave-lengths, so that $\log(l/\lambda)$ is comparable with, or greater than, $\log(\lambda/a)$.

First we shall dispose of case (a). Here as can be seen from the study of Eq. (27) the distinction between $f_p(z)$ and $\bar{f}_p$ is important only at distances from the ends of the wire which are comparable with the wire radius, and small compared

---

[11] E. C. Stoner, *Magnetism and Matter*, p. 39.

with the wave-length. However, because of the fact that the current vanishes at the ends of the wire, the substitution of $\bar{f}_p$ for $f_p$ leads to no serious error when we remember to include the sum of the resonant and forced terms rather than to consider them separately. Individually they would show abrupt discontinuities near the end of the wire if we assumed them inversely proportional to $f_p(z)$. Because $\log(l/\lambda)$ is by hypothesis small compared to $\log(\lambda/a)$ in case (a), the value of $f_p$ is roughly the same for the forced and resonant members, as can be seen from Eqs. (27) and (30). Hence for estimating the orders of errors, we can replace $f_f$, $f_r$ by a common expression $f$, and the point of our argument is that in the "end zone," i.e., the region a small fraction of a wave-length from the end of the wire, the difference between the expressions

$$f^{-1}(D_f \cos qz + D_r \cos\beta z)$$
and
$$\bar{f}^{-1}(D_f \cos qz + D_r \cos\beta z) \qquad (44)$$

is of minor importance. The reason for this statement is that at the ends of the wire the terms $D_f \cos qz$ and $D_r \cos\beta z$ nearly cancel. The cancellation, it was pointed out in Part I, is not complete, but still it is enough to insure that throughout the end zone either form of expression (44), and hence the difference between the two forms, is only of the order $I(0)/\Omega$ (with $\Omega \sim \log(\lambda/a)$). This is sufficient accuracy since our solution only pretends to be accurate to the order $I(0)$. Here $I(0)$ denotes the order of magnitude of the current down most of the wire (not necessarily at the center) and is of the order $\omega E/\beta^2\Omega$ away from resonance and $\omega E/\beta^2$ at resonance. This argument can be stated somewhat more exactly as follows: If we iterate (1) or (25) to obtain a second approximation to the current, the difference as compared with (24) is

$$\Delta I = f(z)^{-1}\Big\{ -\int r^{-1}[I_0(\xi) - I_0(z)] \cos\beta r d\xi$$
$$+ i\int r^{-1} I_0(\xi) \sin\beta r d\xi$$
$$+ (i\omega \cos\varphi/\beta^2 \sin\theta) E_0 e^{iqz}$$
$$+ A_1 \cos\beta z + A_2 \sin\beta z \Big\} - I_0(z), \qquad (45)$$

where $I_0(z)$ denotes the first approximation to the current, given by (24) and where the significant part of $f(z)$ is given by (26). Because of the boundary condition (31) one finds that $I_0(l)$ is of the order $1/\Omega$ compared with $I_0(0)$. The factor in braces in (45) vanishes at the ends of the wire, and reaches values of the order of $I_0(0)\bar{f}$ only when one moves a wave-length or so away from the ends. Except at the very ends of the wire, the distinction between $f$ and $\bar{f}$ disappears, so that the factor in braces in Eq. (45) is nearly the same as $f(z)I_0(z)$. Thus $\Delta I$ is everywhere of the order $I_0(0)/\Omega$. Similarly one can show that each successive approximation gives a contribution of the order $1/\Omega$ times the preceding one. Our procedure is hence convergent.

In case (b), however, the situation is a different one. The value of $f_p(z)$ is essentially different for the forced and resonant parts of the solution, as $f_f$ has the structure $A \log(\lambda/a) + B$, while $f_r$ has the form $C \log(\lambda/a) + D \log(l/\lambda) + E$ and so increases without limit when the wire contains more and more wave-lengths, whereas in (a) we assumed that $\log(l/\lambda)$ was small compared with $\log(\lambda/a)$. In the end zone the correction $\Delta I$ furnished by the second approximation still is of the order $I_0(0)/\Omega$, but becomes of the order $I_0(0)$ at distances from the end of the wire comparable with the wave-length, but large compared with the radius of the wire and small compared with its total length. Since the correction is comparable in order with the first approximation for part of the wire, it may seem that all semblance of convergence is lost. However, we believe that the resulting error is not serious. In the first place, the difficulty arises only in connection with the resonant rather than forced part of the solution, as the extra term $D \log(l/\lambda)$ occurs only in connection with the former, and in the latter most of the contribution comes from the "close-in" rather than the "far-away" portion since

$$\int_\epsilon^\infty \cos q\xi \frac{e^{-i\beta r}}{r} d\xi$$

converges. Hence our forced member is reasonably well determined, and for long wires this is the main effect. Furthermore, the fact that the integral varies only slowly with position on the wire, together with the fact that most of the radiation does not come from too near the end zone, means that the amplitude of even our resonant term is probably not in error by a factor 2 or so until $\log(l/\lambda)$ becomes large rather than merely comparable with $\log(\lambda/a)$. When $\log(l/\lambda)$ is large compared with $\log(\lambda/a)$, the resonant terms are unimportant compared with the forced, and so any quantitative percentage error in the resonant part of the theory in this region is inconsequential as long as it gives resonant terms which are small compared with the forced. In other words, our theory behave correctly in the limit of infinitely long wires, where it agrees with Chu. Before reaching the limit, but by the time $\log(l/\lambda)$ becomes comparable with $\log(\lambda/a)$, the difference between the heights of the peaks and valleys is small compared with the elevations of the valleys, so that accuracy in this difference is unimportant. We might add that practically we are not concerned, anyway, with wires so long that $\log(l/\lambda)$ is comparable with $\log(\lambda/a)$, so that case (b) is of rather academic interest.

In closing, it is perhaps well to make a few remarks comparing our procedure in method (B) to that of Hallén[4] (or King and Harrison[3]), of Gray,[5] and of King and Middleton.[6] In each instance one is concerned with obtaining a first approximation to an integral of the form $\int r^{-1}\cos k\xi e^{-i\beta r}d\xi$ with the singularity at the origin avoided by integrating down the center of the wire, so that $r^2=(\xi-z)^2+a^2$. Hallén approximates the integral by $\cos kz\int r^{-1}d\xi$. When the upper limit of integration is infinity, the integral $\int r^{-1}\cos k\xi e^{-i\beta r}d\xi$ converges when $k\neq\beta$, whereas $\int r^{-1}d\xi$ does not. Thus, it is clear that the Hallén procedure is not appropriate for long wires. This difficulty is avoided by Miss Gray, who approximates the integral by

$$\cos kz\int r^{-1}e^{-i\beta r}d\xi.$$

This is not an unreasonable step when $k\neq\beta$, as then both the original and substituted integrals are convergent. Hence the Gray scheme is a fairly suitable one for the forced part of the solution, as here $k\neq\beta$. In fact, for the broadside case $k=0$ which figures particularly prominently in the forced part of the scattering, the original form of the integral does not differ from that substituted by Miss Gray. In this particular case, there is no essential distinction between the Gray scheme of approximation and our own which resembles that of King and Middleton. (We might mention parenthetically, that the Gray procedure was not intended specifically for the problem of the receiving antenna, to the forced part of which it applies quite well, but rather for the resonant radiating antenna, to which the same difficulties apply as in the resonant portion which we now discuss. On the other hand, it was intended primarily for the half wave antenna, which is so short that the proper behavior, as $l$ approaches infinity, is irrelevant. Even for the half-wave case, the method of King and Middleton is somewhat more accurate than that of Miss Gray.) Certainly, the type of substitution employed by Miss Gray is not a happy choice for the resonant terms if the wire is so long as to contain a great number of wave-lengths. The difficulty is that the resonant part of the problem has $k=\beta$, so that $\int r^{-1}\cos k\xi e^{-i\beta r}d\xi$ diverges (through to half the Hallén value) when the limit is infinite, whereas the integral substituted by Gray converges. Correspondingly, the amplitude of the resonant member, which by (43) is inversely proportional to the substituted integral, remains finite if the Gray approximation is applied to an infinitely long wire, whereas actually it should go to zero. This difficulty is avoided in our type of procedure which is essentially the same as that of King and Middleton and consists in replacing $\int r^{-1}\cos k\xi e^{-i\beta r}d\xi$ by $\cos kz\int r^{-1}\cos k(z-\xi)e^{-i\beta r}d\xi$ in first approximation. (Cf. Eq. (27) or (42).) The factor multiplying $\cos kz$ is then an integral which converges for infinitely long wires for the forced part of the solution and diverges for the resonant part, whereas Hallén has divergence for both portions, and Gray has convergence. Correspondingly in the limit of very long wires, the forced amplitude remains finite, approaching the Chu value asymptotically, and the resonant member approaches zero as should be the case, whereas the first approximation of Hallén would make both amplitudes tend to zero, and that of Gray would make them both stay finite.

# X-Ray Scattering at Small Angles by Finely-Divided Solids. I. General Approximate Theory and Applications*

C. G. Shull** and L. C. Roess
*Beacon Research Laboratory of The Texas Company, Beacon, New York*

(Received August 29, 1946)

The general approximate theory of x-ray scattering at small angles by finely divided solids is reviewed. It is assumed that the x-ray scattering data can be interpreted in terms of a particle size distribution, with particle-to-particle scattering negligible. Scattered intensity curves for Maxwellian, Gaussian, and rectangular size distributions are calculated, and several procedures for obtaining a size distribution from the experimental scattering data are described. Details of the experimental technique are given, and evidence is presented to show that appreciable error can result if crystal-monochromated radiation is not used. General procedures are given for applying corrections to the experimental data for the slit geometry used. Scattering data, mass distribution curves, and average particle sizes are presented for amorphous silica gels and for crystalline oxides of alumina, nickel and iron. These data are shown to correlate well with the results of crystal size measurements by x-ray diffraction line broadening and with specific surface measurements. The average particle size values determined from the experimental data are shown to have considerable significance, at least for the type of materials considered here.

## INTRODUCTION

The usefulness of x-ray scattering at small angles in the study of finely-divided solids has been demonstrated by many investigators.[1] Carbon blacks have been studied by Warren and his collaborators,[1,2] by Guinier,[3] and by Krishnamurti,[1] cellulose in various forms by Kratky and his collaborators,[4] by Hosemann[5] and by Guinier,[3] while Guinier[3,6] has applied the method to ovalbumine, rubber, colloidal silver, ramie, and the hardening of alloys. Catalyst studies have been reported by Guinier,[3] Elkin, Shull, and Roess,[7] and Jellinek and Fankuchen.[8]

The interpretation of the x-ray intensity scattered at small angles is by no means unambiguous.[9] At least two specialized points of view have been emphasized in the literature: (1) the Guinier[1] and Hosemann[1] theory, which emphasizes particle form factor scattering and assumes that particle-to-particle scattering can be neglected, and (2) Kratky's view[4,10] that, at least for fibrous materials, such as cellulose, a statistical distribution of Bragg reflection planes is responsible for the observed scattering.

Basically it is the distribution of electron density which produces the scattering, and therefore nothing more than this distribution, if that much, can be obtained without ambiguity from the x-ray data. An average electron density, the Patterson distribution,[11] can be calculated, at least in principle, from the observed x-ray data. Such an average electron density having been determined, its interpretation in terms of more

---

* Much of the material in this paper was presented by C. G. Shull at two Gibson Island conferences: the ASXRED meeting in 1944 and the AAAS Catalyst Session in 1945.

** Now at Clinton Laboratories, Oak Ridge, Tennessee.

[1] J. A. Gray and W. H. Zinn, Can. J. Research 2, 291–293 (1930); P. Krishnamurti, Ind. J. Phys. 5, 473–488 (1930); H. Mark and J. Hengstenberg, in H. Mark, *Physik und Chemie der Zellulose*, Verlagsbuchhandlung, Julius Springer (Berlin, 1932), p. 139; S. B. Hendricks, Zeits. f. Krist. 83, 503–504 (1932); B. E. Warren, J. Chem. Phys. 2, 551–555 (1934); A. Guinier, Comptes rendus 206, 1641 (1938); O. Kratky, Naturwiss. 26, 94 (1938); R. Hosemann, Zeits. f. Physik 113, 751–768 (1939); The ASXRED *Bibliography on X-Ray Small Angle Scattering* (1946) has additional references.

[2] B. E. Warren, Phys. Rev. 49, 885 (1936); J. Biscoe and B. E. Warren, J. App. Phys. 13, 364–371 (1942).

[3] A. Guinier, Ann. de physique 12, 161–237 (1939); H. Brusset, J. Devaux, and A. Guinier, Comptes rendus 216, 152–154 (1943).

[4] O. Kratky and F. Schoszberger, Zeits. f. physik. Chemie B39, 145–154 (1938); O. Kratky, Naturwiss. 30, 542–543 (1942); O. Kratky, A. Sekora, and R. Treer, Zeits. f. Elektrochemie 48, 587–601 (1942); O. Kratky and A. Wurster, Zeits. f. Elektrochemie 50, 249–255 (1944).

[5] R. Hosemann, Zeits. f. Physik 114, 133–169 (1939); Zeits. f. Elektrochemie 46, 535–555 (1940).

[6] A. Guinier, Proc. Phys. Soc. 57, 310–324 (1945).

[7] P. B. Elkin, C. G. Shull, and L. C. Roess, Ind. Eng. Chem. 37, 327–331 (1945).

[8] M. H. Jellinek and I. Fankuchen, Ind. Eng. Chem. 37, 158–164 (1945); Ind. Eng. Chem. Anal. Ed. 18, 172–175 (1946) (with E. Solomon).

[9] An excellent discussion is given in A. Guinier, J. Chim. Phys. 40, 133–150 (1943). See also R. Hosemann, Zeits. f. Elektrochemie 46, 535 (1940).

[10] Kratky, Zeits. f. Elektrochemie 46, 556 (1940), discussion of Hosemann's paper.

[11] A. L. Patterson, Zeits. f. Krist. A90, 517–542 (1935). D. Wrinch, *Fourier Transforms and Structure Factors* (American Society for X-Ray and Electron Diffraction, 1946), p. 12.

directly useful physical pictures, such as particle size distributions, offers considerable difficulty, and requires that further assumptions be made.

The most useful method of interpreting the small angle x-ray data may well depend upon the nature of the materials being studied. It has seemed to us that the particle size (or mass) distributions resulting from Hosemann's theory, which assumes particle-to-particle scattering to be negligible, are of great value in forming physical pictures of the solid materials which we have studied (silica gels, silica-alumina catalysts, alumina gels, and alumina-base catalysts). We have, therefore, consistently used this method of interpreting the data, and have endeavored to find evidence supporting it from investigations of these materials by entirely independent means, such as physical adsorption and the broadening of ordinary x-ray diffraction lines.

Once this assumption of negligible particle-to-particle scattering is made, the procedure becomes straightforward. The angular distributions of x-ray intensity scattered at small angles by a series of particle mass distributions obtained by varying two parameters in several types of distribution functions have been calculated. The observed data are compared with the calculated curves, after taking into account any necessary corrections for the geometry of the apparatus, and the mass distribution giving the best fit is determined. Spherical particle shape is assumed in all of the work reported here, although calculations have been made[12] for randomly oriented spheroids also.

The justification for these assumptions can only be found in the usefulness of the particle size distributions obtained and in the agreement between predictions based on them and data obtained independently. We believe that on these grounds the assumptions have been shown to be justified for the materials we have studied.

It is possible to invert the x-ray data to give the mass distribution function directly.[13] This procedure was discovered too late for use in presenting our data. It is unlikely that significantly different results would be obtained for the materials studied here.

## GENERAL THEORY OF X-RAY SCATTERING BY A DISTRIBUTION OF RANDOMLY ARRANGED PARTICLES

For purposes of mathematical analysis we consider a scattering specimen which contains small particles arranged in a random fashion. The intensity of x-radiation scattered at a small angle $\epsilon$ by a single particle is given by

$$I(\xi)=I_e\rho_e^2V^2S(R,\xi), \tag{1}$$

where

$$\xi=(2\pi/\lambda)\epsilon, \tag{2}$$

$I_e$ is the Thomson scattering factor for a single electron and is essentially constant for the angles to be considered, $\rho_e$ is the electron density and $V$ the volume of the particle, and $S(R, \xi)$ is the scattering function characteristic of the particle size $R$ and its geometrical shape. The scattering function $S$ is related to the particle structure as the atomic form factor is related to the structure of an atom.

When many particles $N$ are present, as is always the case experimentally, and if their contributions to the scattered intensity are independent of each other, then obviously the scattered intensity is given by

$$I(\xi)=I_eN\rho_e^2V^2S(R,\xi). \tag{3}$$

This is the equation used by Guinier[3] in his early studies.

If a distribution of particle sizes is envisaged, as Hosemann[1] has suggested,

$$I(\xi)=I_e\rho_e^2\int_0^\infty N(R)V^2(R)S(R,\xi)dR \tag{4}$$

where $N(R)dR$ is the total number of particles having sizes in the range $R$, $dR$. This integral equation can be put into a somewhat more usable form by introduction of a mass distribution function and by expressing the particle volume in terms of the particle size. Thus

$$I(\xi)=KI_e\rho_e^2\int_0^\infty M(R)R^3S(R,\xi)dR, \tag{5}$$

with $M(R)dR$ representing the total mass of particles in the size range $R$, $dR$ and $K$ being a constant proportional to the total mass.

---

[12] L. C. Roess and C. G. Shull, J. App. Phys. 18, 308 (1946).

[13] L. C. Roess, J. Chem. Phys. 14, 695–697 (1946).

TABLE I. Relationship between average particle size and parameters in Maxwellian type of distribution.

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $\bar{r}/r_0$ | 0.227 | 0.693 | 1.183 | 1.677 | 2.176 | 2.674 |

In proceeding further, it is necessary to assume a particular particle shape in order to permit evaluation of the scattering function $S$. The scattering function for the general case of spheroidal particles (which includes spherical, rod-shaped, and plate-shaped particles) has been evaluated in an exact form and will be discussed in the second paper[12] of this series. Guinier[3] has introduced an approximate scattering function

$$S_0(R,\xi)=\exp\left[-\frac{r^2}{3}\xi^2\right], \qquad (6)$$

which agrees exactly at very small angles with the exact scattering function for spherical particles. Deviations are obtained, however, at the larger angles and the effect of these on the interpretation of experimental data will also be discussed in the companion paper. In Eq. (6), Guinier has defined $r$ as the radius of gyration (defined as in the usual mechanical case) of the particle. However, the introduction of the radius of gyration is not of sufficient generality to make the Guinier function a good approximation for very asymmetrical particles, as Guinier[3] has shown.[14]

The integral Eq. (5) along with the Guinier scattering function has been used by Hosemann,[5] by Elkin, Shull, and Roess,[7] and by Jellinek, Solomon, and Fankuchen[8] in their studies. Since, in general, not very much is known in advance about the particle shape or whether there is a preferred shape, it would seem that the use of the Guinier scattering function would be as significant as any. A comparison of the results obtained by interpreting a typical set of experimental data with both the Guinier scattering function and the exact scattering function for spherical particles has shown agreement within about 20 percent for the calculated average particle size. In the present paper, the experimental data have all been interpreted by use of the approximate scattering function and the resultant particle size distributions described in terms of a particle diameter.

## Methods of Data Interpretation

In applying Eq. (5) to the interpretation of experimental data, it is necessary to perform a mathematical inversion of the data in order to obtain the desired particle size distribution. A direct inversion procedure making use of the exact scattering function for spheroidal particles has been introduced by Roess,[13] but as mentioned earlier this was discovered too late for application to the present experimental data. Several indirect methods are available, however, which provide a simple and easily-performed procedure for obtaining the particle size distribution, and these have been used in the present study. Basically these indirect methods involve a matching of the experimental scattering curve with either one or a suitable combination of standard scattering curves which have been calculated in advance for selected distributions of particle size. Three types of distributions have been studied: Maxwellian, Gaussian, and rectangular, and these will be considered individually.

### Maxwellian Type of Particle Size Distribution

Hosemann[1] has suggested that the particle size distribution might be represented by an expression of the type $r^n\exp[-(r/r_0)^2]$, where $n$ and $r_0$ are constants whose values are to be adjusted to the experimental data. Following this idea, we let

$$M(r)=\frac{2}{r_0^{n+1}\Gamma\left(\frac{n+1}{2}\right)}r^n\exp[-r^2/r_0^2], \qquad (7)$$

where the coefficient is that which gives a total mass of unity. The average particle size $\bar{r}$, which is defined as that size which divides the mass distribution in half, has been calculated for this distribution in terms of the parameters $r_0$ and $n$, and the resulting values are listed in Table I. It is interesting that $\bar{r}/r_0$ varies almost exactly linearly with $n$ for $n \geqslant 1$.

If Eq. (7) and the Guinier scattering function,

---

[14] See also section on Numerical Results in reference 12.

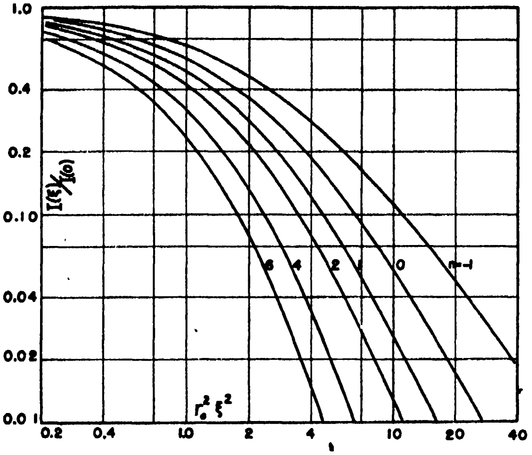


FIG. 1. Calculated scattering curves for Maxwellian type of particle size distribution.

Eq. (6), are inserted into Eq. (5), we obtain

$$I(\xi)=KI_e\rho_e^2\frac{2}{r_0^{n+1}\Gamma\left(\frac{n+1}{2}\right)}$$

$$\times\int_0^\infty r^{n+3}\exp[-(\tfrac{1}{3}\xi^2+r_0^{-2})r^2]dr, \quad (8)$$

and this integrates directly to

$$I(\xi)=\frac{2KI_e\rho_e^2\Gamma\left(\frac{n+4}{2}\right)r_0^3}{\Gamma\left(\frac{n+1}{2}\right)}$$

$$\times[\tfrac{1}{3}r_0^2\xi^2+1]^{-[(n+4)/2]}. \quad (9)$$

The requirement of finite total mass in the distributions of Eq. (7) restricts values of $n$ to those greater than $-1$.

Equation (9) represents the intensity distribution to be expected if the particle size distribution has the form of Eq. (7). Conversely if Eq. (9) can be made to fit the experimental data by adjustment of its parameters, then the particle size distribution is obtainable directly from Eq. (7). The adjustment of the parameters can be conveniently done by either of two graphical methods, (a) a curve matching and resolution process, and (b) a curve shifting procedure.

The former procedure makes use of the *standard scattering curves* shown in Fig. 1. These have been calculated[15] according to Eq. (9). If the experimental scattering curve (plotted as $\log I$ *versus* $\log\xi^2$) can be matched to one of these standard scattering curves by suitable displacement of the two graphs along horizontal and vertical axes, the particle size distribution is represented by Eq. (7) with the parameter $n$ determined by the selected standard curve and $r_0$ determined by comparing the location of the $\xi^2$ axis on the experimental graph in relation to the $r_0^2\xi^2$ axis of the standard curve graph. If, however, the experimental data do not match any of the standard curves it is necessary to resolve the experimental curve into the sum (or difference) of two or more component curves each of which is a good match for one of the standard curves. The parameters of the component curves will be determined in the same fashion as in the earlier one-component match and the relative amplitudes of the distributions will be determined by the relative intensities of the component scattering curves. It has been found generally that one standard curve will represent the experimental data, but even when additional component curves have been found necessary the whole resolution process can be carried out in less than one-half hour.

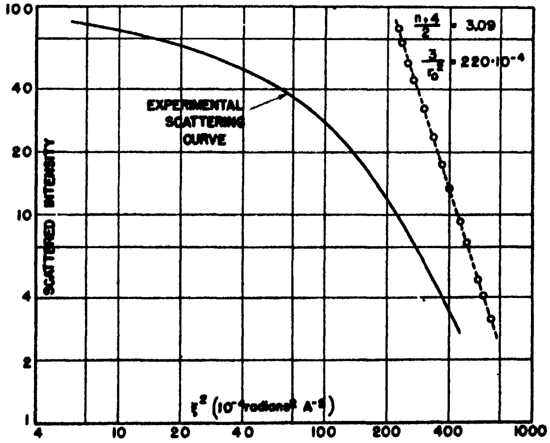


FIG. 2. Typical experimental scattering curve for alumina gel illustrating the shifting procedure for determining the particle size distribution parameters.

---

[15] The calculated values for the scattered intensity produced by various particle size distributions have been tabulated and are available from the authors upon request.

The second method for determining the particle size distribution does not involve the use of the standard curves but is nevertheless quite similar to the former method. Equation (9) can be put into a logarithmic form

$$\log I(\xi)=\log K_1-\left(\frac{n+4}{2}\right)\log[\xi^2+3/r_0^2], \quad (10)$$

with $K_1$ a constant. From this equation it can be seen that a graph of the experimental intensity *versus* $(\xi^2+3/r_0^2)$ on a double logarithmic scale should show a linear relationship between these variables. Moreover the slope of the line through the experimental data would be $-[(n+4)/2]$ and hence $n$ would be determined. It is not possible to plot the above variables directly, since $r_0$ is also an undetermined parameter, but the data can be plotted simply as $I$ *versus* $\xi^2$ (again on double logarithmic paper) and then various values of $3/r_0^2$ added to $\xi^2$ until a straight line is obtained. Figure 2 illustrates the use of this shifting procedure. The experimental data are represented by the solid line and upon adding a fixed quantity ($220\cdot 10^{-4}$ radians$^2$ $A^{-2}$) to the abscissa of the experimental values, the circled points are seen to agree closely with a straight line. The parameters $n$ and $r_0$ are immediately available and the particle size distribution can be represented by Eq. (7). In cases where the shifting process does not lead to a single straight line, the experimental curve can be broken into component curves each of which can be shifted into a straight line. Additional terms are therefore necessary in describing the true particle size distribution. Again this process can be performed in a short time.

## Gaussian Distribution of Particle Size

Let the particle size distribution be represented by

$$M(r)=\frac{2}{\sqrt{\pi}}\frac{\beta}{r_0}\frac{1}{1+H(\beta)}\exp\left[-\frac{\beta^2}{r_0^2}(r-r_0)^2\right], \quad (11)$$

where the coefficient has been selected to make the total mass unity. Substitution of Eqs. (11) and (6) into Eq. (5) gives upon integration

$$I(\xi)=\frac{KI_e\rho_e^2}{\sqrt{\pi}}\frac{r_0^3}{\beta^{\frac{1}{2}}[1+H(\beta)]}D(\beta, r_0\xi) \quad (12)$$

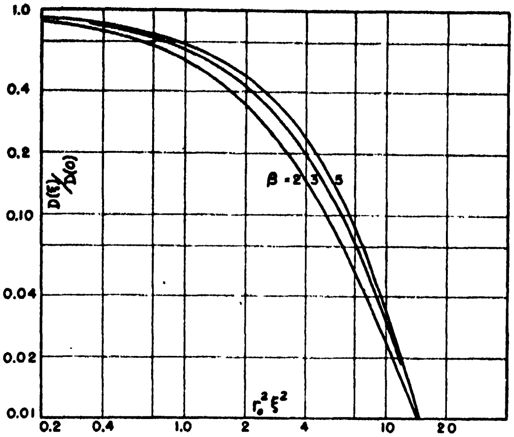


FIG. 3. Calculated scattering curves for Gaussian type of particle size distribution.

with

$$D(\beta, r_0\xi)=\beta^{-13/2}\tau^4\bigg\{(1+\tau^2)\exp(-\beta^2)$$

$$+\frac{\sqrt{\pi}}{2}\tau\exp(\tau^2-\beta^2)[2\tau^2+3][1+H(\tau)]\bigg\}, \quad (13)$$

where

$$\tau^2=\frac{3\beta^4}{3\beta^2+r_0^2\xi^2}$$

and

$$H(x)=\frac{2}{\sqrt{\pi}}\int_0^x \exp(-u^2)du.$$

From Eq. (12) it is seen that the angular dependence of the scattered intensity is wholly within the function $D$. This function has been evaluated for various values of $\beta$, and a typical family of curves is shown in Fig. 3. It should be mentioned that the parameter $\beta$ in the distribution given by Eq. (11) determines the sharpness of the distribution, whereas $r_0$ is the average particle size.

The curves given in Fig. 3 can be used in the same fashion as those of Fig. 1 in interpreting experimental data. If the data can be matched to one of the curves, one expression of the type given by Eq. (11) suffices to describe the particle size distribution. If a simple matching cannot be performed it is necessary to perform the resolution process as outlined for the Maxwellian type of distribution.

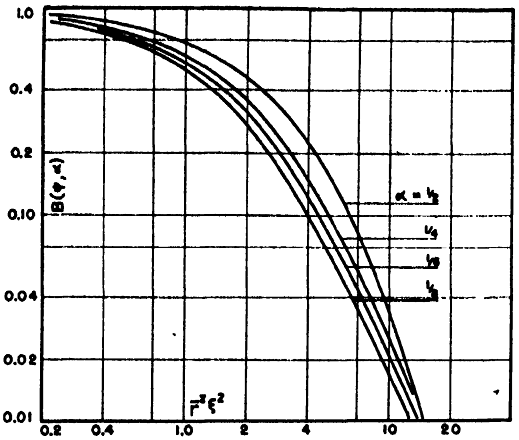


FIG. 4. Calculated scattering curves for rectangular type of particle size distribution.

## Rectangular Distribution of Particle Size

Let the particle size distribution be represented by

$$M(r)=\frac{1}{r_2-r_1}\quad\text{for}\quad r_1\leqslant r\leqslant r_2$$
$$=0\ \text{otherwise.}\qquad(14)$$

Equation (5) now integrates into

$$I(\xi)=\frac{2KI_e\rho_e^2(1-\alpha^4)}{(1-\alpha)(1+\alpha)^3}\bar{r}^3B(\psi,\alpha),\qquad(15)$$

where

$$B(\psi,\alpha)=\frac{2}{(1-\alpha^4)\psi^4}\{(1+\alpha^2\psi^2)\exp(-\alpha^2\psi^2)-(1+\psi^2)\exp(-\psi^2)\}\qquad(16)$$

with

$$\alpha=\frac{r_1}{r_2},\quad \bar{r}=\frac{r_1+r_2}{2}\quad\text{and}\quad \psi^2=\frac{4}{3}\frac{\bar{r}^2}{(1+\alpha)^2}\xi^2.\qquad(17)$$

In Eq. (17), $\bar{r}$ is the average particle size of the rectangular distribution. The angular variation of the scattered intensity is given by the function $B$, and this is graphed in Fig. 4. Curves of this type may be used to evaluate the particle size distribution in a fashion similar to that outlined above.

## Comparison of the Scattering Curves Calculated for Different Distributions

The standard scattering curves which have been given in the preceding sections are all seen to possess three distinct features: (a) a flat section at the very small angles, (b) a curved region (the knee of the curve) at intermediate angles and (c) a steep, linear section at the larger angles. Because of the difficulty of making absolute intensity measurements, the interpretation of the experimental data is based on the shape of the experimental scattering curve, rather than on its absolute level. Since a fit between the observed curve and one of a family of calculated curves is made by effectively multiplying the ordinates and abscissae of the former by arbitrary factors (displacement in the vertical and horizontal directions until a match is found), it is seen that no significant match can be found if the experimental curve has been obtained in an angular region which does not include the curve knee.

In comparing the families of curves in Figs. 1, 3, and 4, it is found that some members in one family are very similar in shape to others in another family. Specifically, the curve for $\beta=2$ in Fig. 3 is very similar (over the angular range illustrated) to the curve for $\alpha=1/4$ in Fig. 4. This means that very accurate and extensive experimental data would be required in order to select the correct distribution form. However, the average particle sizes, obtained by matching an experimental curve with both of these calculated curves, agree very closely in value. Hence it appears that the curve matching method may introduce some ambiguity as to the distribution form but does lead to quite definite values of the average particle size for those cases in which the particle geometry is such that the Guinier scattering function can be used.

Figures 5 and 6 illustrate several particle size distribution curves and their appropriate calculated scattering curves. Distributions $A$, $B$, and

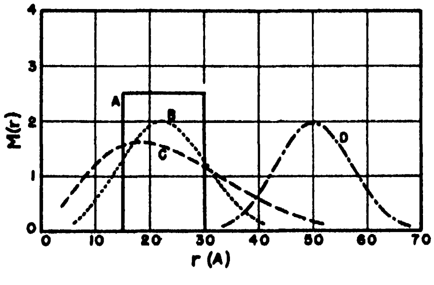


FIG. 5. Various types of particle size distributions.

*C* all have closely the same average particle size. The scattering curves for *A*, *B*, and *C* differ somewhat in the knee curvature with the sharpest distribution (A) showing the most curvature in the scattering curve. Distribution *D* corresponds to a considerably different average particle size and the resultant scattering curve differs markedly from the other curves.

### Experimental Technique

A schematic diagram of the camera used in determining x-ray scattering at small angles is shown in Fig. 7. Basically the camera consists of two parts: (a) a crystal monochromator for insuring that x-rays of only one wave-length pass into the camera, and (b) the scattering chamber which is an evacuated cylinder about 15 cm in diameter and 60 cm in length. Both calcite and rocksalt crystals are used as monochromating crystals: the former when a high degree of collimation is desired and the latter when higher intensity and less collimation is found necessary in the primary x-ray beam. The x-radiation coming from the monochromator passes into the evacuated scattering chamber through a thin beryllium window and then through slits *A* and *B* of respective opening 0.13 by 1.6 and 0.13 by 2.5 mm. These slits serve both to limit the width of the primary beam and to eliminate the weaker component of the $K\alpha$ doublet when a calcite crystal is used. If only one slit is used, the well resolved $K\alpha$ doublet will pass through the camera, and because of the angular separation of its components an unsymmetrical scattering pattern will be obtained. By proper alignment of the calcite crystal and the two slits, a homogeneous primary beam $K\alpha_1$, can be obtained. The half width of this beam after passage through slit *B* has been determined to be 0.10 mm, and at a distance of 50 cm beyond this slit this value has increased to only 0.14 mm. Thus it is seen that very little divergence is present in the beam reflected from the calcite crystal. This collimation is very desirable when the x-ray scattering is to be studied at the very small angles.

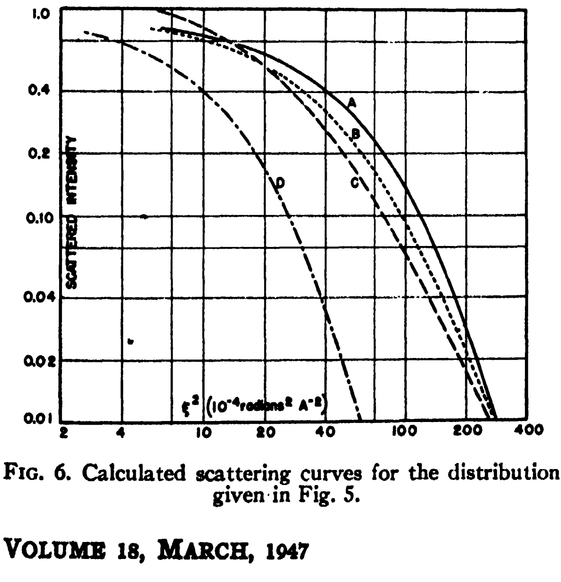


FIG. 6. Calculated scattering curves for the distributions given in Fig. 5.

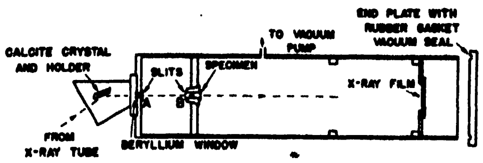


FIG. 7. Schematic diagram of small angle x-ray scattering camera.

The scattering specimen is placed immediately behind the second slit and the scattered radiation recorded on a piece of photographic film located at one of several positions in the scattering chamber. It has been found convenient to pulverize the sample by grinding to a particle size of the order of one micron (much larger than the particles whose size is to be studied by the scattering) and to pack this powder into a rectangular opening about (4 by 2 mm) in a flat metal plate. The thickness of the plate determines the specimen thickness and this can be selected so as to be of optimum value. No binder has been found necessary in this technique. A beam stop of width 0.5 mm is placed 1 cm in front of the film to absorb the primary beam. Specimen to film distances from 60 to over 400 mm have been used, depending upon the range of scattering angles to be investigated. At the conclusion of the main scattering exposure of duration 10 minutes up to 30 hours depending upon the scattering power of the specimen, the beam stop is removed and a short exposure of the primary beam is made. A typical exposure is reproduced in Fig. 8. The image of the primary beam and the scattered radiation on either side of the beam can be seen.

The intensity distribution across the equator (perpendicular bisector of the primary beam

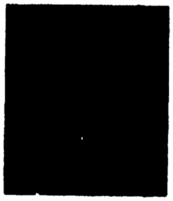


FIG. 8. Typical small angle x-ray scattering photogram.

image) of the film is determined with a microphotometer of scanning area 0.04 by 0.5 mm. Since the scattered x-ray intensity varies over a considerable range (several hundred or a thousand to one) a series of exposures are made with increasing exposure times in order that a large range of photographic density may be covered without exceeding the range in which the x-ray intensity is proportional to the density on any one film. After the individual films are microphotometered, the component intensity curves are combined by graphical matching into a complete scattering curve.

Some tests have been performed in which the use of crystal monochromated radiation is compared with that of filtered and unfiltered x-radiation. Figure 9 shows a series of scattering curves obtained for a sample of alumina with three types of primary radiation. The geometrical conditions of exposure (beam divergence, etc.) were kept unchanged from one exposure to another. It will be seen that intensity errors as large as 50 percent are obtained with unfiltered radiation. Upon filtering out the $K\beta$ radiation component, intensity errors as large as 20 percent are still present, caused by the white radiation band which is transmitted through the filter. The average particle size values as calculated from the filtered and crystal monochromated radiation curves differ by about 25 percent.

It has been mentioned earlier that slit openings, rather than pinholes, were used to define the primary beam. This has been done in order to utilize the increased intensity of scattering available with a slit system. Such technique, however, requires that a correction be applied to the data since the scattered intensity distribution will be altered. Hosemann[1] has considered the case in which the slit openings are infinitely long, but this situation will not be satisfied under the usual experimental conditions. The general theory of slit correction has been worked out and will be outlined in Appendices A and B of this report. As will be shown there, two methods of applying the correction are available and either one may be used.

## Experimental Results

The above procedure of analysis has been applied to several dozen finely divided samples in this laboratory. In general, scattering data were taken over the scattering angle range from about 0.2 to 2.5 degrees. As mentioned above, for the best interpretation of the data it should cover the region of interesting curvature. This region will depend in an approximately inverse fashion on the average particle size: with small particle systems the larger scattering angles are most interesting, while the smaller scattering angles are most profitably studied for systems of large average particle size.

In view of the interpretative assumptions discussed earlier, it is highly desirable to compare the results of x-ray scattering analysis with those obtained by other independent techniques. With crystalline materials the crystal size (as contrasted with the particle size) can be evaluated from studies of the width of x-ray diffraction lines, and for both crystalline and amorphous materials, low temperature gas adsorption can be used for obtaining information on the state of subdivision. Both methods have been used in our studies wherever possible.

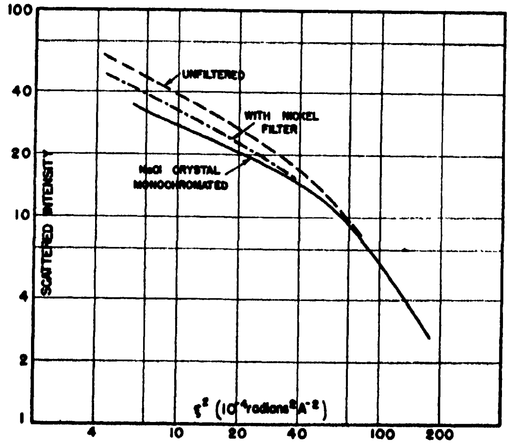


FIG. 9. Comparison of experimental scattering curves obtained with unfiltered, filtered and crystal monochromated primary radiation.

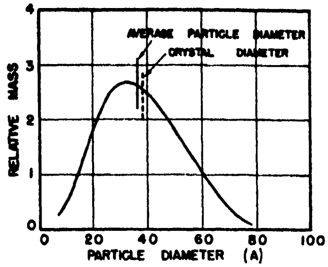


FIG. 10. Particle size distribution obtained for a sample of alumina gel.

In determining the crystal size, the Jones[16] technique of correcting the line broadening for geometrical effects was used. Values of the Scherrer constant used in the crystal size evaluation were those given by Stokes and Wilson[17] and Murdock.[18]

A typical particle size distribution[19] obtained for a sample of alumina gel is shown in Fig. 10. The experimental data and their analysis for this alumina gel have been given in Fig. 2. From the distribution curve, an average particle size can be readily evaluated. For the distribution given in Fig. 10 the average particle size has been determined as 36 angstroms and this has been marked on the distribution curve by the full vertical line. Also shown on the distribution curve by the broken vertical line is the crystal size (38 angstroms) as determined from the diffraction line broadening. The agreement between the crystal and particle sizes is well within the experimental error.

Figure 11 illustrates the correlation which has been obtained between average crystal and average particle sizes. The line drawn represents equivalence of the two values. It is to be emphasized that the particle size may be either equal to or larger than the crystal size, since agglomeration of crystals into larger particles may be present. Crystals larger than particles would not be expected however. The four shaded points to the lower right of the line are believed to signify this agglomeration effect. These four materials were $Al_2O_3$-$MoO_3$ catalysts which had been prepared in a special manner. Independent evidence supporting this hypothesis is obtained from the gas adsorption studies in which the particle size correlated better than the crystal size with the specific surface.

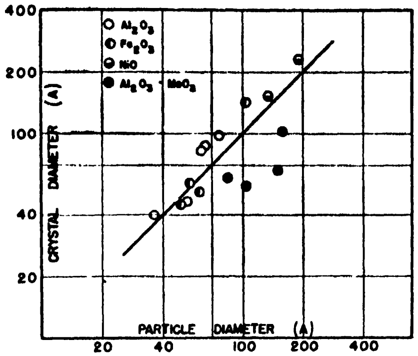


FIG. 11. Correlation between average crystal and particle sizes for a variety of materials. The solid points are believed to represent crystal agglomeration into larger particles.

The unshaded points in Fig. 11 indicate fair agreement between the crystal and particle size. It would appear from the general sequence of the data that the crystal size is a little larger than the particle size, and this may reflect the difference in degree of weighting which the two determinations afford. Jones[16] has shown that the crystal size values will be weighted to the larger sizes and this may account for the observed displacement of values from the line.

Low temperature gas adsorption analysis has been widely used[20] in the study of finely divided solids and this technique provides interesting data for comparison with the x-ray scattering data. One can easily determine the specific surface (surface area per unit mass) and over-all porosity of such solids by this method. It can be shown that the specific surface $S$ of a collection of spherical particles of diameter $D$ is given by

$$S=\frac{6\int_0^\infty \frac{M(D)}{D}dD}{\rho\int_0^\infty M(D)dD} \tag{18}$$

$$=\frac{6}{\rho\bar{D}} \tag{19}$$

---

[16] F. W. Jones, Proc. Roy. Soc. 166, 27 (1938).

[17] A. R. Stokes and A. J. C. Wilson, Proc. Camb. Phil. Soc. 38, 313 (1942) and 40, 197 (1944).

[18] C. C. Murdock, Phys. Rev. 63, 223 (1943).

[19] When using the Guinier exponential scattering function, the particle size distribution is obtained in terms of the radius of gyration of the particle. For spheres, the particle diameter is $2(5/3)^{\frac{1}{2}}$ times the radius of gyration.

[20] P. H. Emmett, Ind. Eng. Chem. 37, 639 (1945).

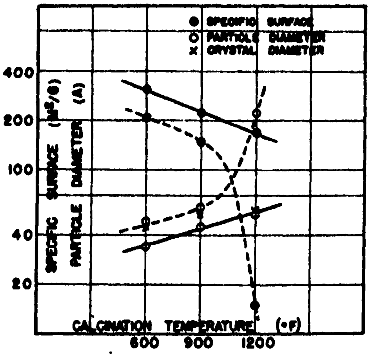


FIG. 12. Average particle size and specific surface variation with calcination temperature for two iron oxide preparations. Solid line—preparation *A*, broken line—preparation *B*.

where $\rho$ is the particle density (not over-all apparent density) and $\bar{D}$ is an average particle size. Thus a linear relationship should exist between $S\rho$ and $1/\bar{D}$.

The specific surface and average particle sizes for a large number of materials have been determined.[21] Typical of the results of such investigations have been those published[7] for various silica-alumina preparations. It was found there that modifications in preparation technique brought about changes in the specific surface of the gel which were followed in all cases by appropriate qualitative changes in the average particle size as determined by small angle x-ray scattering technique.

Figure 12 illustrates additional comparative data obtained by gas adsorption and x-ray scattering analyses. Two preparations of iron oxide were heat treated at the temperatures shown on this graph and the specific surface and average particle size values were obtained. It is seen that for preparation *B* the specific surface falls rapidly and the particle size increases rapidly upon heat treatment about 1000°F. Preparation *A* on the other hand is seen to be quite stable in regard to either measurement at temperatures through 1200°F. Also shown on the graph are several crystal size values and these are seen to agree closely with the particle size determinations.

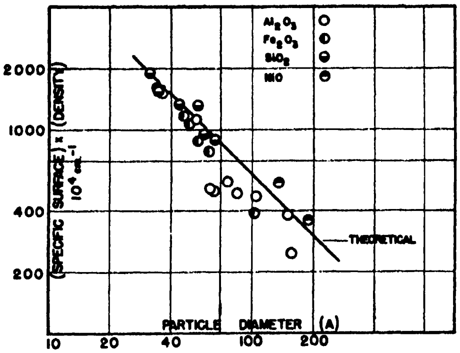


FIG. 13. Correlation between specific surface and particle size for a variety of materials.

Equation (19) predicted a linear relationship between $S\rho$ and $1/\bar{D}$ and Fig. 13 summarizes our data with regard to this. The line shown on this graph is that given by Eq. (19). Considerable deviations are to be noted for individual points but the general trend of the data follows fairly well that to be expected. Data for both amorphous and crystalline materials are represented on this graph.

## ACKNOWLEDGMENT

It is a pleasure to acknowledge the stimulation and encouragement received from Professor B. E. Warren of M.I.T., and several illuminating discussions of interpretational details with Professor A. L. Patterson of Bryn Mawr.

## APPENDIX A. METHOD OF CORRECTING THE SCATTERED INTENSITY DISTRIBUTION OBTAINED WITH A SLIT-SHAPED BEAM TO THAT FOR A PINHOLE BEAM

The experimental data given in the body of this paper were all taken with a slit-shaped primary beam, whereas the theory given is applicable only to data taken for a pinhole beam of very small cross section. Fortunately it is possible to change the experimental intensity distribution into a form equivalent to that which would have been obtained if a pinhole primary beam had been used. The theory of data interpretation given in the body of this paper and the companion paper can then be applied to this modified scattering curve.

A ray diagram of the x-ray optics in the scattering camera is shown in Fig. 14. The x-rays coming from the crystal or target focal line (not shown), in front of the first slit (of length $2l_1$), fan

---

[21] We are indebted to Dr. P. B. Elkin for supplying us with the gas adsorption results.

out from this slit in a vertical plane. Those which pass through the second slit (length $2l_2$), and strike the specimen, are scattered through the scattering angle and may be either (1) recorded on a piece of photographic film placed in a plane perpendicular to the camera axis, or (2) detected by a Geiger counter behind a scanning slit opening. In either case the intensity distribution is determined by means of a scanning slit of finite length ($2y_0$) at different positions along the equator axis. It is obvious that the intensity at any point in the scanning area is a result of the superposition of various intensity elements as scattered through different scattering angles by different scattering elements in the sample.

Consider a specimen element $ds$ which is struck by a ray making an angle $\eta$ with the camera axis. From the camera geometry it is seen that

$$\begin{aligned} d &= \epsilon L, \\ u &= \alpha L, \\ g &= \eta L + s, \end{aligned} \tag{20}$$

since all of the angles under consideration are small (less than 3°). The angle $\alpha$ is characteristic of the position of the scanning slit in the plane of intensity measurement. Since

$$d^2 = u^2 + (g-y)^2,$$

it follows that

$$\begin{aligned} \epsilon^2 &= \alpha^2 + \left(\eta + \frac{s-y}{L}\right)^2, \\ &= \alpha^2 + \sigma^2 \end{aligned} \tag{21}$$

with

$$\sigma = \eta + \frac{s-y}{L}. \tag{22}$$

$g(\alpha)$ will be considered to represent the intensity striking the scanning area with center at angular position $\alpha$, and $I(\epsilon)ds$ the intensity of scattering at an angle $\epsilon$ by a specimen element $ds$. Since $g(\alpha)$ is the summation of the intensity produced by all specimen elements and all primary rays, it follows upon averaging over the scanning area, that

$$g(\alpha) = \frac{1}{2y_0} \int_{-y_0}^{y_0} dy \int_{-l_2}^{l_2} ds \int_{(s-l_1)/L_3}^{(s+l_1)/L_3} I(\epsilon) d\eta. \tag{23}$$

The problem remaining is the determination of $I(\epsilon)$ when $g(\alpha)$ is known. Suppose $I(\epsilon)$ to be represented by

$$I(\epsilon) = \sum_i T_i \exp(-t_i^2 \epsilon^2), \tag{24}$$

with as many terms included in the summation as are necessary to describe the function. It follows from Eqs. (21) and (23) that

$$g(\alpha) = \frac{1}{2y_0} \sum_i T_i \exp(-t_i^2\alpha^2) \int_{-y_0}^{y_0} dy \int_{-l_2}^{l_2} ds \times \int_{(s-l_1)/L_3}^{(s+l_1)/L_3} \exp(-t_i^2\sigma^2) d\eta, \tag{25}$$

$$= \sum_i T_i f(t_i) \exp(-t_i^2\alpha^2), \tag{26}$$

$$= \sum_i T_i' \exp(-t_i^2\alpha^2), \tag{27}$$

with

$$f(t_i) = \frac{1}{2y_0} \int_{-y_0}^{y_0} dy \int_{-l_2}^{l_2} ds \int_{(s-l_1)/L_3}^{(s+l_1)/L_3} \times \exp(-t_i^2\sigma^2) d\eta, \tag{28}$$

and

$$T_i' = T_i f(t_i). \tag{29}$$

The function $f(t_i)$ is independent of $\alpha$ and depends only upon values of $t_i$ and the camera geometry, and the latter remains fixed during the various exposures of one specimen (assuming the specimen-to-detector distance remains unchanged). Equations (24) and (27) permit the desired transformation of $g(\alpha)$ into $I(\epsilon)$ in the following fashion. If $g(\alpha)$ can be represented by a series of exponential terms, each with known values of $t_i$ and $T_i'$, then $I(\epsilon)$ can be represented by the same exponential series but with different term amplitudes, $T_i$.

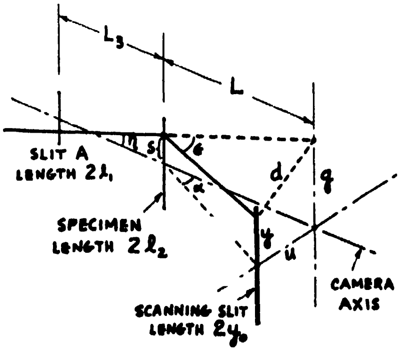


FIG. 14. Ray diagram of the x-ray optics in the scattering camera.

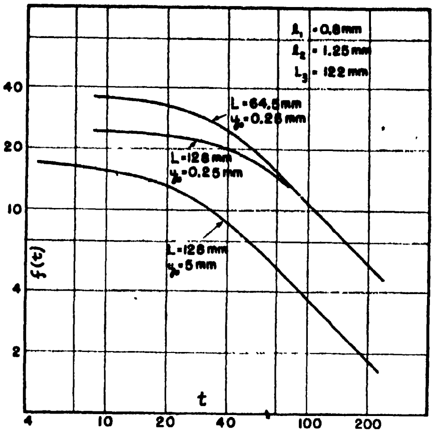


FIG. 15. Slit correction function calculated for various experimental conditions.

The triple integral in Eq. (28) has been integrated with the result that

$$
\begin{aligned}
\frac{4hct_i^3}{L_3}f(t_i)=\frac{\sqrt{\pi}}{2}\{&[1+2t_i^2(a+c)^2]H[t_i(a+c)]\\
&+[1+2t_i^2(c-a)^2]H[t_i(c-a)]\\
&-[1+2t_i^2(b+c)^2]H[t_i(b+c)]\\
&-[1+2t_i^2(c-b)^2]H[t_i(c-b)]\}\\
&+t_i\{(a+c)\exp[-t_i^2(a+c)^2]\\
&+(c-a)\exp[-t_i^2(c-a)^2]\\
&-(b+c)\exp[-t_i^2(b+c)^2]\\
&\quad-(c-b)\exp[-t_i^2(c-b)^2]\},\qquad(30)
\end{aligned}
$$

where

$$
\begin{aligned}
h&=1+\frac{L_3}{L},\\
a&=\frac{hl_2+l_1}{L_3},\qquad(31)\\
b&=\frac{hl_2-l_1}{L_3},\\
&y_0/L_1
\end{aligned}
$$

and

$$H(x)=\frac{2}{\sqrt{\pi}}\int_0^x\exp(-v^2)dv.$$

Equation (30) has been evaluated numerically for various $t_i$ and the function $f(t_i)$ is graphed in Fig. 15 for the experimental conditions generally used in our particular scattering camera. In the present study a microphotometer slit opening of length 0.5 mm was used for scanning across the photographic film and typical curves for this scanning area are given in Fig. 15. If a scanning slit of longer length had been used (as would have been done with a Geiger counter) the values of $f(t_i)$ would have been changed and this is also illustrated in Fig. 15. It is to be emphasized that $f(t_i)$ depends only upon the camera geometry and need be evaluated just once.

The representation of $g(\alpha)$ by a series of exponential terms can be conveniently performed graphically. When the experimental intensity distribution is graphed on log intensity *versus* $\alpha^2$ axes, the curve can be resolved into the sum of a series of straight lines (as is done in analyzing complex radioactive decay curves, for instance). Each straight line corresponds to one of the terms in Eq. (27) and $t_i$ and $T_i'$ can be determined from the slope and intensity axis intercept, respectively, of the straight line. Knowing the values of $f(t_i)$ from Fig. 15, the true scattering curve (for pinhole conditions) can then be reconstructed by use of Eq. (24).

## APPENDIX B. SIMPLIFIED METHOD OF APPLYING THE SLIT CORRECTION WHEN THE EXPONENTIAL SCATTERING FUNCTION IS USED

In Appendix A a general method of correcting the scattered intensity for the slit geometry is given. When the Guinier scattering function is used in interpreting the data, a second and more easily performed procedure is available for taking into account the slit effect. In this procedure the correction is applied to the final particle size distribution rather than to the original intensity data.

The basic integral expression for the intensity of x-ray scattering at an angle $\epsilon$ by a small scattering specimen $ds$ has been given (Eq. (5)) as

$$I(\xi)ds=KI_e\rho_e^2\int_0^\infty M(r)r^3\exp(-\tfrac{1}{3}r^2\xi^2)drds,\quad(32)$$

with $\xi=(2\pi/\lambda)\epsilon$. When many specimen elements and many primary and scattered rays are contributing to the observed intensity, additional integrations must be performed as has been discussed in Appendix A. Applying Eqs. (21) and (23) to (32), the observed intensity distribution is given by

$$g(\alpha)=KI_e\rho_e{}^2\int_0^\infty M(r)r^3$$

$$\times\exp\left(-\frac{4\pi^2}{3\lambda^2}r^2\alpha^2\right)dr\frac{1}{2y_0}\int_{-y_0}^{y_0}dy\int_{-l_2}^{l_2}ds$$

$$\times\int_{(s-l_1)/L_3}^{(s+l_1)/L_3}\exp\left(-\frac{4\pi^2}{3\lambda^2}r^2\sigma^2\right)d\eta, \quad (33)$$

$$=KI_e\rho_e{}^2\int_0^\infty M(r)r^3$$

$$\times\exp\left[-\frac{4\pi^2}{3\lambda^2}r^2\alpha^2\right]drf\left(\frac{2\pi}{\sqrt{3}\lambda}r\right), \quad (34)$$

with $f[(2\pi/\sqrt{3}\lambda)r]$ given by Eq. (28). Let

$$M(r)f\left(\frac{2\pi}{\sqrt{3}\lambda}r\right)=M'(r). \quad (35)$$

It follows that

$$g(\alpha)=KI_e\rho_e{}^2\int_0^\infty M'(r)r^3\exp\left[-\frac{4\pi^2}{3\lambda^2}r^2\alpha^2\right]dr, \quad (36)$$

and this expression has the same form as the original integral expression for pinhole conditions. Thus the function $M'(r)$ can be obtained by procedures given in the body of this paper. Following this, the true particle size distribution can be derived with the use of Eq. (35).

---

## Back Number Prices

ATTENTION is called to the new price scale for single copies and back volumes of this journal as listed on the second cover page. The new price scale was effective March 1, 1947.

The increase is the result of increased handling costs and the fast-diminishing supply of older issues. The increase will make it possible for exhausted issues to be reprinted with no great burden placed on current subscribers for this reprinting.

# X-Ray Scattering at Small Angles by Finely-Divided Solids. II. Exact Theory for Random Distributions of Spheroidal Particles

L. C. ROESS AND C. G. SHULL*
*Beacon Research Laboratory of The Texas Company, Beacon, New York*

(Received August 29, 1946)

The theory of the scattering of x-rays at small angles is given for a continuous distribution in size of randomly spaced and oriented spheroidal particles having arbitrary, but fixed, shape. Families of scattering curves are presented for spheres, and for spheroids ranging in shape from flat disks to long rods. Both Maxwellian and rectangular types of particle mass distribution are used. A fit between the experimental and a calculated scattering curve enables one, under favorable circumstances, to determine the mass distribution in the test sample. However, an unambiguous interpretation of the experimental scattering curve on the basis of the small angle scattering theory is not possible without additional evidence from independent investigations of such quantities as particle shape and sample specific surface.

## INTRODUCTION

IT has been pointed out in the preceding paper[1] that a reasonable, though by no means entirely unambiguous, approach to the interpretation of x-ray scattering at small angles can result from the calculation of the scattering to be expected from assumed mass distributions of particles having a definite shape, followed by an estimation of the actual mass distribution and particle shape by obtaining a fit between the observed scattered intensity curve and one of a series of calculated intensity curves. It is the object of this paper to present the exact theory for the x-ray scattering at small angles by a continuous distribution of randomly spaced spheroidal particles having an arbitrary, but fixed, shape. Scattering by spheres is, of course, a special case of the theory, while the scattering by disks of finite radius but vanishing thickness, and by rods of finite length but vanishing radius form two limiting cases.

## ASSUMPTIONS

It is desirable to state clearly and exactly the assumptions on which the calculations are based:

1. The particles have whatever form of random spacing necessary to produce incoherency in the radiation scattered by individual particles, so that the total intensity is the sum of the individual intensities.
2. The particles have random orientation.
3. The incident x-ray beam is composed of monochromatic and parallel rays, and is of negligibly small cross-sectional extent.
4. The interaction of the scattered radiation with the incident radiation inside the sample to produce an index of refraction is ignored. This effect should be noticeable only at extremely small angles.[2] A closely related assumption is that multiple scattering can be neglected.
5. The effects of finite sample size are ignored. They also should be noticeable only at extremely small angles.
6. The effects of variations on an atomic scale of the electron density distribution inside a particle are ignored.
7. The particles all have the same shape.[3]
8. Absorption of x-radiation inside a particle can be neglected.

## GENERAL THEORY

The general equation for the scattered intensity $I(\xi)$ is[4]

$$I(\xi)=AI_e\rho_e^2\int_0^\infty V^2(R)N(R)\langle S(R,\xi)\rangle_{Av}dR, \quad (1)$$

$$=AI_e\frac{\rho_e^2}{\rho}\int_0^\infty V(R)M(R)\langle S(R,\xi)\rangle_{Av}dR, \quad (2)$$

with

$$\xi=\frac{4\pi}{\lambda}\sin\theta, \quad (3)$$

$\lambda$ = wave-length,
$2\theta=\epsilon$ = scattering angle,

---

* Now at Clinton Laboratories, Oak Ridge, Tennessee.

[1] C. G. Shull and L. C. Roess, J. App. Phys. 18, 295 (1947).

[2] R. von Nardroff, Phys. Rev. 28, 240–246 (1926).

[3] The mathematical methods used in this paper have also been applied to calculate the scattered intensity for particle distributions varying in both shape and size.

[4] R. Hosemann, Zeits. f. Physik 113, 751–768 (1939). See also Eq. (4) of reference 1.

$I_e$ = Thomson scattering factor for a single electron,[5]
$A$ = absorption factor,[6]
$\rho_e$ = average electron density of particle, electrons per cm$^3$,
$\rho$ = mass density of individual particles, gram cm$^{-3}$.

$V(R)$ is the volume of a particle characterized by the size parameter $R$, $\langle S(R, \xi)\rangle_{\mathrm{Av}}$ is the scattering function (the shape function[7] or form factor for the particle averaged over-all orientations), and $N(R)dR$ and $M(R)dR$ are, respectively, the number and the mass of the particles having size parameters in the range $R$, $dR$.

## SHAPE FUNCTIONS FOR SPHEROIDS

Patterson[7] and Guinier[8] have shown that the shape function for a spheroid having principal axes $R$, $R$, $vR$ in a system of Cartesian coordinates $(x, y, z)$ is

$$S(R\xi, v) = S(u) = 9\left(\frac{\sin u - u\cos u}{u^3}\right)^2, \tag{4}$$

with

$$u = R\xi(\sin^2\vartheta + v^2\cos^2\vartheta)^{\frac{1}{2}}. \tag{5}$$

In Eq. (5), $\vartheta$ is the angle between the $z$ axis and the difference $\mathbf{s}-\mathbf{s}_0$ between the unit vectors $\mathbf{s}_0$ and $\mathbf{s}$ which specify the direction of the incident and scattered beams, respectively.

The effect of random orientation of the spheroids can be obtained most simply by considering the spheroid fixed, and averaging over-all directions of the vector $\mathbf{s}-\mathbf{s}_0$. This is equivalent to integrating $S$ over the surface of a sphere, and gives

$$\langle S(R\xi, v)\rangle_{\mathrm{Av}} = \frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi} S[R\xi(\sin^2\vartheta + v^2\cos^2\vartheta)^{\frac{1}{2}}]\sin\vartheta d\vartheta d\varphi,$$

or, putting $x = \cos\vartheta$,

$$\langle S(R\xi, v)\rangle_{\mathrm{Av}} = \int_0^1 S[R\xi(1+(v^2-1)x^2)^{\frac{1}{2}}]dx. \tag{6}$$

It follows directly from formulas given by Watson[9] that Eq. (4) may be written

$$S(u) = \frac{9\pi}{2}\frac{J_{3/2}{}^2(u)}{u^3}, \tag{7}$$

$$= {}_1F_2(2; 4, (5/2); -u^2). \tag{8}$$

The generalized hypergeometric function is defined[10] by

$$\begin{aligned}{}_pF_q(\alpha_1, \alpha_2, \cdots\alpha_p; \beta_1, \beta_2, \cdots\beta_q; z) \\ = \sum_{m=0}^{\infty}\frac{(\alpha_1)_m(\alpha_2)_m\cdots(\alpha_p)_m z^m}{m!(\beta_1)_m(\beta_2)_m\cdots(\beta_q)_m},\end{aligned} \tag{9}$$

with

$$(\alpha)_0 = 1; \quad (\alpha)_m = \alpha(\alpha+1)\cdots(\alpha+m-1). \tag{10}$$

Obviously ${}_pF_q$ is an integral function when $q \geqslant p$.

Substituting Eq. (8) in Eq. (6), and integrating term by term, using[11]

$$\int_0^1 [1+(v^2-1)x^2]^m dx = {}_2F_1(-m, \tfrac{1}{2}; \tfrac{3}{2}; 1-v^2), \tag{11}$$

one obtains

$$\langle S(R\xi, v)\rangle_{\mathrm{Av}} = \sum_{m=0}^{\infty}\frac{(2)_m(-R^2\xi^2)^m}{m!(4)_m(5/2)_m} \times {}_2F_1(-m, \tfrac{1}{2}; \tfrac{3}{2}; 1-v^2), \tag{12}$$

$$= {}_1G_2(2; 4, (5/2); -R^2\xi^2, v), \tag{13}$$

where we define

$$\begin{aligned}{}_pG_q(\alpha_1, \alpha_2, \cdots\alpha_p; \beta_1, \beta_2, \cdots\beta_q; z, v) \\ = \sum_{m=0}^{\infty}\frac{(\alpha_1)_m(\alpha_2)_m\cdots(\alpha_p)_m z^m}{m!(\beta_1)_m(\beta_2)_m\cdots(\beta_q)_m} \\ \times {}_2F_1(-m, \tfrac{1}{2}; \tfrac{3}{2}; 1-v^2).\end{aligned} \tag{14}$$

For spherical particles, $v=1$, ${}_2F_1(-m, \frac{1}{2}; \frac{3}{2}; 0) = 1$, and $\langle S(R\xi, 1)\rangle_{\mathrm{Av}} = S(R\xi)$.

In the limit $v \to 0$

$${}_2F_1(-m, \tfrac{1}{2}; \tfrac{3}{2}; 1-v^2) \to {}_2F_1(-m, \tfrac{1}{2}; \tfrac{3}{2}; 1) = \frac{(1)_m}{(\frac{3}{2})_n}, \tag{15}$$

so that

$$\langle S(R\xi, 0)\rangle_{\mathrm{Av}} = {}_2F_3(1, 2; 4, (5/2), \tfrac{3}{2}; -R^2\xi^2). \tag{16}$$

---

[5] Because of the small angular range involved, the state of polarization of the incident beam has negligible effect on the angular variation of intensity.

[6] The absorption may be assumed independent of angle for the small angles under discussion.

[7] A. L. Patterson, Phys. Rev. 56, 972–977 (1939).

[8] A. Guinier, Thèses (Masson et Cie, Paris, 1939), or Ann. de Physique 12, 161–237 (1939).

[9] G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, Cambridge, 1922), pp. 54 and 147.

[10] For example, reference 9, p. 100.

[11] Note that Eq. (11) is valid for any value of $m$ if $v \neq 0$.

This is the scattering function for a disk of radius $R$ and vanishingly small thickness.

If we let $v \to \infty$, but simultaneously $R \to 0$ in such a way that $Rv \to L$, the spheroid approaches a rod of length $2L$ and vanishingly small radius. The scattering function for such a rod may be obtained most simply from Eq. (6).

$$\lim_{\substack{R\to 0\\ v\to\infty\\ Rv\to L}} \langle S(R\xi, v)\rangle_{\mathrm{Av}} = \langle S_\infty(L\xi)\rangle_{\mathrm{Av}} = \int_0^1 S(L\xi x)dx \tag{17}$$

$$= {}_2F_3(\tfrac{1}{2}, 2; 4, (5/2); \tfrac{3}{2}; -L^2\xi^2), \tag{18}$$

since

$$\frac{1}{2m+1} = \frac{(\frac{1}{2})_m}{(\frac{3}{2})_m}. \tag{19}$$

By direct integration, using Eq. (4), it can readily be shown that

$$\langle S_\infty(x)\rangle_{\mathrm{Av}} = \frac{6}{5}\frac{Si(2x)}{x} - \frac{3}{2x^4} - \frac{9}{10x^6} + \frac{3\cos 2x}{5x^2}\left(1 - \frac{1}{2x^2} + \frac{3}{2x^6}\right) + \frac{3}{10}\frac{\sin 2x}{x^3}\left(1 + \frac{6}{x^2}\right), \tag{20}$$

in which

$$Si(x) = \int_0^x \frac{\sin t}{t} dt. \tag{21}$$

A similar closed expression for $\langle S(R\xi, 0)\rangle_{\mathrm{Av}}$ has not been found.

### MASS DISTRIBUTION FUNCTIONS

Hosemann[4] introduced the Maxwellian type of mass distribution function[12]

$$M(R) = B\left(\frac{R}{R_0}\right)^n e^{-(R/R_0)^2};$$

$$B = \frac{2}{R_0\Gamma\left(\frac{n+1}{2}\right)}, \tag{22}$$

which has a maximum at

$$R_{\max} = (n/2)^{\frac{1}{2}} R_0, \tag{23}$$

and showed that a wide variety of distribution forms could be obtained by varying the parameters $n$ and $R_0$. Consideration of the more general function

$$M(R) = B\left(\frac{R}{R_0}\right)^n e^{-(R/R_0)^m} \tag{24}$$

has shown that no advantage with respect to variety of the mass distribution form is gained by choosing $m \neq 2$, while considerable simplicity is lost in evaluating the integral in Eq. (2).

It must be pointed out here that, since the incident beam has a finite cross section, the effective scattering specimen has a finite volume, so that

$$M(R) = 0, \quad R > R',$$

where $R'$ is of the order of the beam radius. A mass distribution function expressed by

$$\left.\begin{aligned} M(R) &= BR^n(R'-R)^\mu, & 0 \leq R \leq R' \\ &= 0, & R > R' \end{aligned}\right\} \tag{25}$$

in which $\mu$ is very large, has been tried and found to give results essentially the same as the Maxwellian type. The increased simplicity of the analysis using the latter type is thus sufficient to justify its use, despite the infinite range of $R$.

### INTENSITY SCATTERED BY MAXWELLIAN MASS DISTRIBUTIONS

Since for a spheroid

$$V(R) = \frac{4\pi}{3} vR^3, \tag{26}$$

substitution of Eqs. (26), (22), and (13) in Eq. (2), and term by term integration gives, using Eq. (14),

$$I(\xi, v) = K \sum_{m=0}^{\infty} \frac{(2)_m \left(\frac{n+4}{2}\right)_m (-R_0{}^2\xi^2)^m}{m!(4)_m(5/2)_m} \times {}_2F_1(-m, \tfrac{1}{2}; \tfrac{3}{2}; 1-v^2), \tag{27}$$

$$= K\,{}_2G_2\left(2, \frac{n+4}{2}; 4, (5/2); -R_0{}^2\xi^2, v\right), \tag{28}$$

with

$$K = AI_e \frac{4\pi v R_0{}^3 \rho_e{}^2}{3\rho} \frac{\Gamma\left(\frac{n+4}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}. \tag{29}$$

---

[12] With the value of $B$ given in Eq. (22), the total mass is unity.

Equation (27) is the formal solution of the scattering problem. It is worth noting that the parameter $R_0$ enters directly as a scale factor, while the parameter $n$ affects the shape of the intensity function.

The series in Eq. (27) converges for all values of $\xi$, $R_0$, $n$, and $v$. However, numerical calculation by direct summation of the series becomes impractical for large values of $R_0\xi$ or $R_0\xi v$. In these ranges asymptotic series must be used for the calculation; the form of these series depends upon the value of $v$. Three ranges of $v$ are important:

$$v \ll 1; \quad v \cong 1; \quad \text{and} \quad v \gg 1.$$

In each range, a suitable form of ${}_2F_1(-m, \tfrac{1}{2}; \tfrac{3}{2}; 1-v^2)$ must be used.[13]

It can be shown that Eq. (27) can be written

$$I(\xi, v) = \frac{K}{(1-v^2)^{\frac{1}{2}}} \times \left\{ {}_3F_3\left(1, 2, \frac{n+4}{2}; 4, \frac{5}{2}, \frac{3}{2}; -R_0{}^2\xi^2\right) - \frac{v^2}{2} \sum_{m=0}^{\infty} \frac{\left(\frac{n+4}{2}\right)_m (-R_0{}^2\xi^2 v^2)^m}{(4)_m (5/2)_m} \times {}_2F_1(\tfrac{1}{2}, m+1; m+2; v^2) \right\}. \tag{30}$$

Equation (30), while valid for all values of $v$, is useful for numerical calculation when $v \ll 1$. For large $R_0\xi$ an asymptotic expansion[13] of the first term is required.

Equation (27) can also be written

$$I(\xi, v) = K \sum_{m=0}^{\infty} \frac{(2)_m \left(\frac{n+4}{2}\right)_m (-R_0{}^2\xi^2)^m}{m!(4)_m(5/2)_m} \times {}_3F_3\left(2+m, \frac{n+4}{2}+m, \frac{1}{2}; 4+m, \frac{5}{2}+m, \frac{3}{2}; -R_0{}^2\xi^2(v^2-1)\right). \tag{31}$$

Equation (31) is also valid for all values of $v$ and, when used with the asymptotic expansion[13] for the ${}_3F_3$ terms, is suitable for numerical calculation when $v \gg 1$.

When $v \cong 1$, the series in Eq. (27) suffices for numerical calculation when used with its asymptotic expansion[13] for large $R_0\xi$.

## INTENSITY SCATTERED BY RECTANGULAR MASS DISTRIBUTIONS

If the mass distribution is defined as follows

$$\left.\begin{aligned} M(R) &= \frac{1}{R_2 - R_1}, \quad R_1 \leqslant R \leqslant R_2 \\ &= 0, \qquad \text{otherwise} \end{aligned}\right\} \tag{32}$$

so that it is normalized to unit mass, the scattered intensity given by Eq. (2) becomes, using Eq. (26),

$$I(\xi, v) = \frac{K'}{R_2 - R_1} \times \{R_2{}^4 {}_2G_3(2, 2; 4, 3, (5/2); -R_2{}^2\xi^2, v) - R_1{}^4 {}_2G_3(2, 2; 4, 3, (5/2); -R_1{}^2\xi^2, v)\}, \tag{33}$$

with

$$K' = AI_e \frac{4\pi v \rho_e{}^2}{3\rho}. \tag{34}$$

If we put $\alpha = R_1/R_2$, Eq. (33) can be written

$$I(\xi, v) = \frac{K'R_2{}^3}{1-\alpha} \times \sum_{m=0}^{\infty} \frac{(2)_m(2)_m\, {}_2F_1(-m, \tfrac{1}{2}; \tfrac{3}{2}; 1-v^2) \times (-R_2{}^2\xi^2)^m(1-\alpha^{4+2m})}{m!(4)_m(3)_m(5/2)_m}. \tag{35}$$

The series (divided by $1-\alpha^4$) has been calculated for $v=1$ and $\alpha = 0$, 0.5, 0.9, and 1.0. (When $\alpha = 1$, the intensity is that produced by a group of spheres all having the same radius $R_2$, and is given by Eq. (8) with $u = R_2\xi$.)

## NUMERICAL RESULTS

Calculated scattered intensity curves[14] are plotted on a double logarithmic scale on Figs. 1 to

---

[13] A summary of the formulas giving the relevant asymptotic series, as well as a discussion of the procedure used in obtaining them, has been prepared and is available from the authors on request.

[14] Numerical values have been tabulated, and are available from the authors upon request.

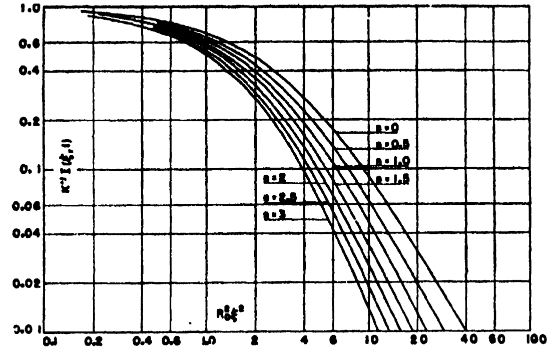


FIG. 1. Calculated intensity scattered by Maxwellian distributions of spheres, $v=1$ (Eq. (27)).

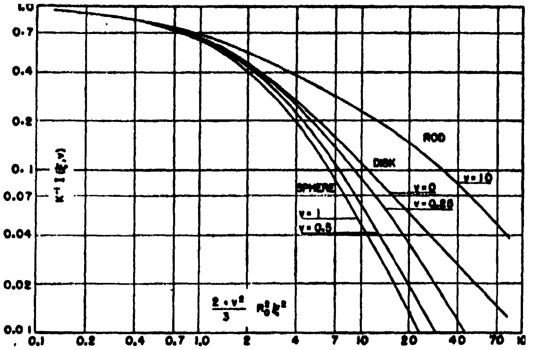


FIG. 2. Calculated intensity scattered by the same Maxwellian distribution of spheroids having various shapes.

5. Figure 1 shows the effect of varying the parameter $n$ (which controls the form of the mass distribution) on the shape of the angular distribution of the intensity scattered by a Maxwellian distribution of spheres. The increased curvature observed as $n$ increases is associated with an increased sharpness in the mass distribution. As is shown by the asymptotic expansions, all curves for $n>0$[15] approach the same slope $(-2)$ as $R_0^2\xi^2$ becomes large.

Figure 2 shows the effect of particle shape ($v=0$, $\frac{1}{4}$, $\frac{1}{2}$, 1, 10) on the angular distribution of the intensity scattered by a mass distribution having a fixed form ($n=1$). The abscissa are chosen proportional to the radius of gyration[16] of the particle in order to make the curves merge at small values of $R_0\xi$. Distinct differences are apparent between the disk-shaped ($v=0$) and rod-shaped ($v=10$) particles. The smaller slopes at large $R_0\xi$ values are predicted by the asymptotic expansions: for $v=0$ the limiting slope is $-1$ and for $v$ very large the limiting slope is $-\frac{1}{2}$.

Figures 3 and 4 show the effect of varying $n$ on the shape of the angular distribution of the intensity scattered by oblate spheroids having the fixed shape parameters $v=0$ and $v=\frac{1}{4}$, respectively.

Figure 5 shows the calculated angular distribution of scattered intensity produced by rectangular mass distributions (cf. Eq. (32)) having various values of $\alpha=R_1/R_2$. The curvature is considerably greater than that associated with the Maxwellian distributions. Also, although it is not shown on the plot, the scattered intensity shows a series of maxima and minima, for $\alpha=1$, 0.9, and 0.5 at $R_2\xi$ values greater than those for which the curves are drawn.

It may be shown by superposition that the curve for $v=\frac{1}{2}$, $n=1$ on Fig. 2 has very nearly the same shape as that for $v=1$, $n=\frac{1}{2}$ on Fig. 1, while the curves $v=1$, $n=0$ and $v=\frac{1}{4}$, $n=5$ form a similar isomorphic pair. These examples show that when both particle shape and mass distribution form are varied, near coincidences can occur, rendering the interpretation in the scattered intensity curves very uncertain. Obviously, if a distribution in particle shape were also considered, the possible ambiguity would become even greater. Other independent evidence must be taken into account in deciding between the various possibilities. In general, a mass distribution

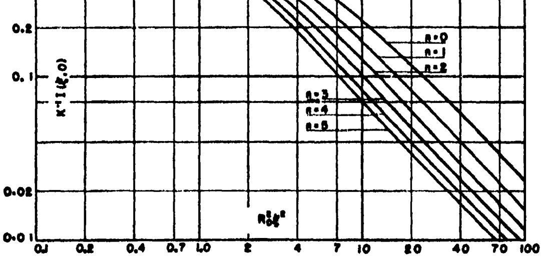


FIG. 3. Calculated intensity scattered by Maxwellian distributions of thin disks, $v=0$.

---

[15] When $n=0$ the leading term in the asymptotic expansion is proportional to $\frac{\ln R_0\xi}{(R_0\xi)^4}$.

[16] $R_g=\left(\frac{2+v^2}{5}\right)^{\frac{1}{2}}R$, cf. A. Guinier, Ann. de Physique 12, 161-237 (1939).

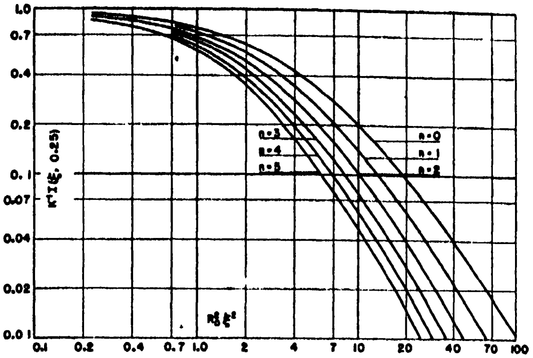


FIG. 4. Calculated intensity scattered by Maxwellian distributions of oblate spheroids, $v=0.25$.

having a small value of $n$ is improbable, so that the ambiguity of the above examples could be at least partially resolved by this argument. If a crystalline material is under investigation, independent information on particle shape can be obtained from x-ray line width measurements.[17]

In each of the examples cited above only one scattering curve is involved. When the radius of gyration of the average particle in each of the two mass distributions calculable from each scattering curve is computed, it is found that those of the spheroidal particles are appreciably greater than those of the spherical particles. This shows that when a given intensity curve is interpretable in terms of two or more distributions of differently shaped particles, the resulting average radii of gyration are by no means identical, as would be required, for example, by the approximate theory using the Guinier scattering function.[1]

The matching of the scattered intensity curves with the experimental data has been discussed in the previous paper.[1]

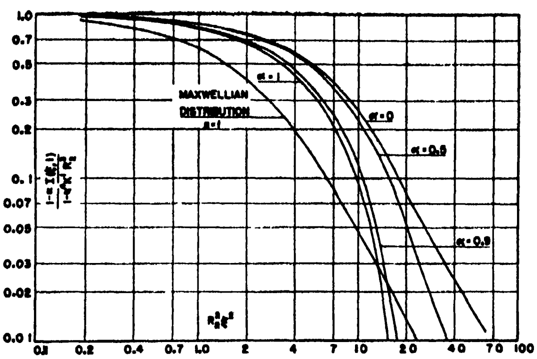


FIG. 5. Calculated intensity scattered by rectangular distributions of spheres (Eq. (35)). $\alpha=R_1/R_2$. The intensity fluctuations at large values of $R_2^2\xi^2$ are not shown. The intensity scattered by a Maxwellian distribution of spheres is also plotted for comparison.

## ERRORS DUE TO USE OF THE GUINIER SCATTERING FUNCTION

The exact scattering function for randomly oriented particles of fixed size and shape has a primary peak centered at zero scattering angle and a series of secondary peaks at larger angles whose amplitudes decrease as some negative power of the angle. In contrast, the Guinier function used in the approximate theory given in the preceding paper[1] has a primary peak only, and decreases exponentially at larger angles. This neglect of the secondary peaks leads to appreciable errors in the calculated scattered intensity at relatively large angles.[18] However, these errors are small for $R_0\xi$ values up to 1.5 or 2, and since the region enclosing the knee of curve[1] rarely extends to $R_0\xi$ values greater than 2.5 to 3, no really serious errors of interpretation result from the use of the Guinier function in place of the corresponding exact scattering function, when the particles are reasonably symmetrical in shape.

---

[17] For the crystalline materials discussed in reference 1, no evidence for shape asymmetry has been found from line width measurements.

[18] This may be illustrated by the asymptotic proportionality of the Guinier function scattered intensity to $(r_0\xi)^{-(n+4)}$, while the exact function for spheres gives $(R_0\xi)^{-4}$ for all $n>0$.

# Excess Noise in Cavity Magnetrons

ROBERT L. SPROULL*
*RCA Laboratories, Princeton, N. J.*†

(Received October 17, 1946)

The magnetron oscillator of some frequency-modulation radars also furnishes the "local oscillator" excitation for the crystal mixer of the receiver. Excess noise generation by the magnetron was observed to reduce greatly the receiver sensitivity. This noise exhibited a strong dependence upon anode voltage and current, and changed with time in a perplexing manner. After many experiments, a hypothesis of the cause of excess noise was developed, and further experiments confirmed this hypothesis. The noise is thought to be caused by ionization of atoms of the cathode oxide coating, which atoms are removed from the cathode by electron bombardment. In order to reduce the generation of excess noise, and still preserve the advantages of an oxide-coated cathode, a special shape of cathode has been developed. The coated regions of this cathode are sheltered from electron bombardment, and the noise is much reduced. Use of cathodes of this general type should also produce magnetrons with longer operating lives than present tubes.

THE magnetron oscillator in some[1] frequency-modulation, continuous-wave radars also furnishes the "local oscillator" excitation of the crystal mixer of the receiver. It was observed that the magnetron oscillations were frequently modulated with large amounts of noise, thereby impairing the sensitivity of the receiver. The cause of this excessive noise modulation (many times ordinary shot noise) was made the subject of a special investigation, the results of which are reported here. Excess noise is not noticed in the usual use of a magnetron as a transmitter tube, since the signal-to-noise ratio is at least 60 db. However, the noise phenomena may

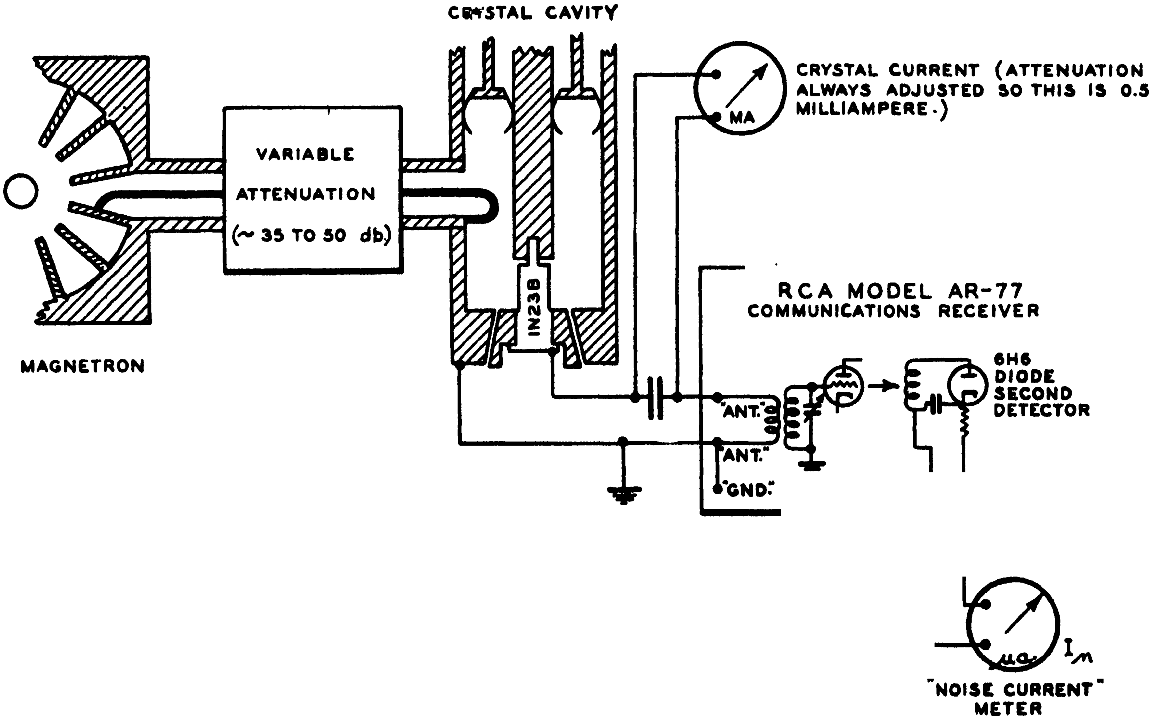

CRYSTAL CAVITY
CRYSTAL CURRENT (ATTENUATION ALWAYS ADJUSTED SO THIS IS 0.5 MILLIAMPERE.)
MA
VARIABLE ATTENUATION (~ 35 TO 50 db)
1N23B
MAGNETRON
RCA MODEL AR-77 COMMUNICATIONS RECEIVER
"ANT."
"ANT."
"GND."
6H6 DIODE SECOND DETECTOR
μa
$I_n$
"NOISE CURRENT" METER

FIG. 1. Noise comparison apparatus.

---

* Now at Cornell University, Ithaca, N. Y.

† This paper is based on work done for the U. S. Navy under Contract NXsa 35042.

[1] Such as the AN/APQ-19, developed in this laboratory.

increase our understanding of magnetron operation, are important in special applications of magnetrons, and are of interest in any development of receiving tubes utilizing magnetic fields.

## I. DESCRIPTION OF THE NOISE PHENOMENA

Figure 1 shows the experimental arrangement used to detect noise modulation and to estimate its magnitude. Since the magnitude of the oscillating field in the crystal cavity was kept constant by varying the attenuation between the magnetron and this cavity, the receiver detector current "$I_n$" was roughly inversely proportional to the signal-to-noise ratio of the magnetron oscillations.

Figure 2 shows the "noise current" $I_n$ as a function of magnetron d.c. anode voltage ($V_a$) for several anode currents ($I_a$) in a 4000 megacycle, continuous-wave magnetron[2] developed in this laboratory. To obtain such a plot, the magnetic field was varied to produce different values of $V_a$ at constant $I_a$. It is apparent that at low voltages the total noise current is substantially constant, and in this region the contribution of the magnetron noise to the total receiver noise is small. At some voltage, which we shall call $V_{ao}$, a relatively sharp break in the curve appears, and above this voltage the magnetron noise dominates noise from other sources and becomes hundreds or thousands of times as large as at low voltages. This region ($V_a > V_{ao}$) of excess noise contained the ordinary operating point of the oscillator. Phenomena very similar to those portrayed in Fig. 2 were also observed in a magnetron[3] of quite different design.

Curves such as those of Fig. 2 are not reproducible in detail. The "peaks" and "valleys" of the curves cannot be repeated even a few minutes after they are observed. The most useful feature for comparing noise phenomena in different tubes is the voltage $V_{ao}$. Even this parameter changes with time, however, and may have quite different values in two tubes constructed (as nearly as possible) identically. Several tubes must, therefore, be constructed in order to be certain of the dependence or lack of dependence of noise upon some one feature of the design. Figure 2 is representative of the early 4000 megacycle tubes and exhibits the increase in noise with decrease in anode current which was observed with all tubes.

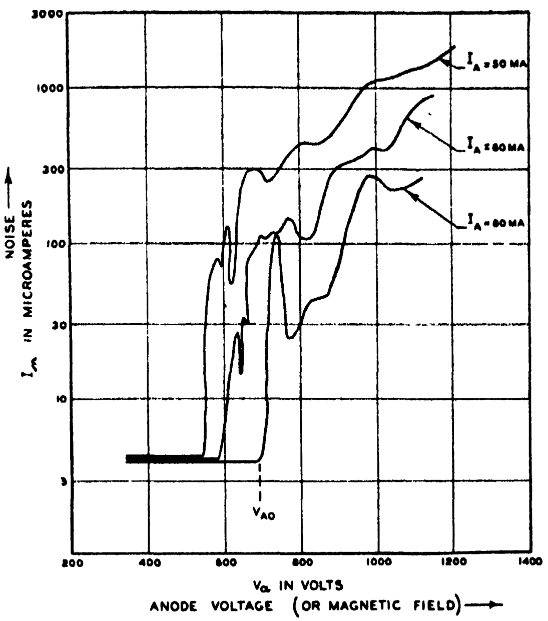


FIG. 2. Noise in early 4000 megacycle magnetrons. At constant anode current, anode voltage and magnetic field are approximately proportional. Therefore the abscissa may be considered (except for units) as either voltage or magnetic field strength.

After vigorous oscillation in these magnetrons has begun at a sufficiently high anode voltage, the cathode heater power may be removed. Back-bombardment of the cathode by electrons which have received energy from the oscillating field produces secondary electrons and warms the cathode sufficiently to produce some thermionic electrons. The excess noise was invariably less when the ordinary heater power was removed. All of the measurements and correlations reported here refer to operation with no applied heater power.

The receiver (Fig. 1) was ordinarily tuned to pass modulation components in the frequency range 1496 to 1504 kilocycles, but other frequency regions were also explored. In general the noise was less the higher the receiver frequency,

---

[2] G. R. Kilgore, C. Shulman, and J. Kurshan, "A Frequency-Modulated Magnetron for Super-High-Frequencies," Proc. I.R.E. (to be published).

[3] J. S. Donal, Jr., R. R. Bush, C. L. Cuccia, and H. R. Hegbar, "A One-Kilowatt Frequency-Modulated Magnetron for 900 Megacycles," Proc. I.R.E. (to be published).

and the excess noise was quite small at a receiver frequency of 30 megacycles.

## II. CORRELATIONS OF NOISE WITH MAGNETRON PROPERTIES

Many tubes were constructed with slightly different features from the ordinary magnetron in order to test the dependence of noise upon different aspects of the design. In each case correlation of noise properties with the changes made was attempted. Correlations were also attempted between noise properties and operating conditions of the tube.

The high-frequency load presented to the magnetron by the antenna and transmission system changed the details of the excess noise, but any standing wave ratios and phases which produced reasonably stable operation with moderate efficiency produced about the same values of $V_{ao}$. The presence or absence of discontinuities in the straps of the magnetron resonator, the presence or absence of "hats" on the ends of the cathode produced no consistent changes in noise. The geometry and potentials of electrodes near the ends of the cathode affected the noise only very slightly. Small changes in anode diameter produced no effect. A variation of almost a factor of two in cathode size produced marked changes in efficiency and stability; but the noise was little changed except with very small or very large diameters, and in these cases the noise was greater than with intermediate diameters.

These and other experiments were used to show that several more or less obvious hypotheses of the causes of excess noise were not tenable. For example, any axial oscillations (such as observed in magnetron structures many years ago by Megaw[4]) should be strongly affected by end plate potentials, and if the noise were related to such oscillations it should be a function of these potentials. Other hypotheses were concerned with the resonant modes of the resonator and with the space-charge and electric field distributions in the cathode-anode space. The results of the studies with various cathode and anode diameters, the fact that the excess noise was observed in a magnetron of quite different construction, and the occurrence of noise over a wide range of anode voltages and currents, made such hypotheses of the cause of noise very unlikely.

Three successful correlations emerged from this work, but none gave immediately a clear indication of the cause of the noise. These correlations will be briefly described here. The first was a correlation of noise with low efficiency accompanied by high back-bombardment of the magnetron cathode. Tubes which exhibited considerably lower efficiency than others of similar construction were usually very noisy. Experimental tubes were constructed with windows which allowed observation of the cathode during operation, and this correlation was verified in such tubes. This correlation is probably related to the reduction in noise as the normal cathode heater power is removed; in both cases high noise accompanies high cathode temperature.

The second correlation was between excess noise and the length of the cathode sleeve which was coated with oxides. It was found that if the emitting length were confined to less than two-thirds of the length of the anode, the voltage ($V_{ao}$) at which noise began could be materially increased. It was not certain that the ends of the cathode did not emit electrons, but the carbonate spray was carefully cleaned from all except the central region of the cathode sleeve in the hope that no emission would occur in the end regions. The suggestion was that the r-f and d.c. fields were distorted there, and it might be desirable to prevent the interaction of electrons with such fields.

The third correlation was between noise and the thermionic emission properties of the cathodes. The critical voltage ($V_{ao}$) was used as a measure of the former, and the heater current required to permit a thermionic current of 50 milliamperes at 100 volts (no magnetic field) was used as a measure of the latter. The correlation coefficient[5] of $V_{ao}$ with this current was 0.79 for thirteen experiments with ten tubes. This shows a definite relation, with low noise accompanying low thermionic activity. This correlation was strengthened by observations on a single tube during the early part of its life, when both the

---

[4] E. C. S. Megaw, "Fundamental Characteristics of Electronic Oscillations," Nature 137, 189 (1936).

[5] H. L. Rietz, *Mathematical Statistics*, Open Court (Chicago, 1927), p. 82.

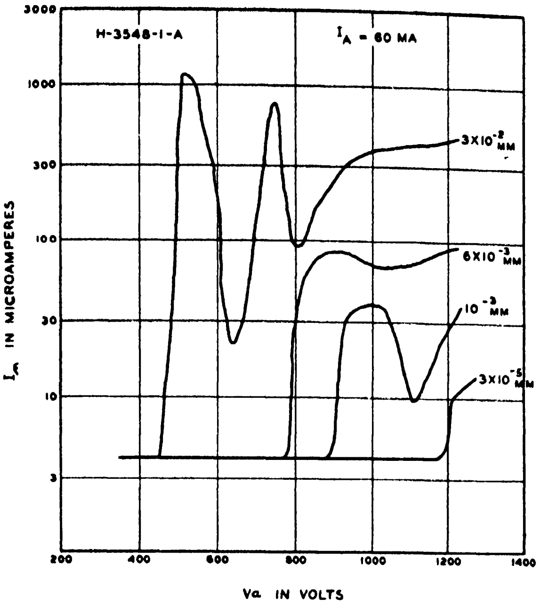


FIG. 3. Noise as a function of pressure. This tube had a helical tungsten filament in place of the usual oxide-coated cathode.

cathode activity and the noise were changing. Magnetrons were constructed with strontium oxide coatings on the cathodes, since the emission from SrO is known[6] to be considerably less than that from the SrO-BaO mixtures ordinarily used. These tubes showed high values of heater current for 50 milliamperes thermionic current, but the noise properties did not differ appreciably from other tubes. This experiment indicated that noise was not directly related to the cathode's thermionic activity, but that for cathodes of the same composition, high noise accompanied high activity.

## III. IONIZATION NOISE

One of the first hypotheses of the cause of the excess noise was that it originated in the ionization of gas atoms in the interaction space of the magnetron. To test this, several tubes were continuously pumped while noise data was taken, and neon gas was admitted to vary the pressure. No variation of noise with pressure was observed for pressures less than about $5\times10^{-3}$ mm of mercury. At pressures greater than this, the noise increased and the efficiency decreased rapidly; this behavior was not surprising since at $10^{-2}$ mm a glow discharge was beginning. The pressures could not be measured inside the magnetron itself, but they were measured in an ionization gauge connected to the magnetron.

Several magnetrons were constructed with tungsten helices in place of the usual oxide-coated cathodes. Two of these were continuously pumped, and noise was observed as a function of pressure. Figure 3 represents the behavior of these tubes. Unlike the results with oxide-cathodes, the data here show a continuous increase in $V_{ao}$ (that is, decrease in noise) as the pressure decreases. For pressures less than $10^{-6}$ mm, $V_{ao}$ was greater than 1200 volts for tungsten filament tubes. But lowering the pressure below $10^{-3}$ mm did not change the noise in oxide-coated cathode tubes, and even at $10^{-6}$ mm $V_{ao}$ was no larger than 800 or 900 volts. If the cause of the noise is the same in the tubes with these two types of cathodes, it must be that there is a condensable gas in the oxide-coated cathode magnetrons which is present in the anode-cathode space but which is not measured by external pressure gauges.

Evidence to support this interpretation came from quite different experiments. Observations upon tubes with windows revealed a luminescence in the anode-cathode space or upon the cathode surface. Spectrographic investigation showed that the line spectra of barium and strontium (the metallic constituents of the cathode coating) were represented by their most persistent lines. Since the lines were quite sharp, it was apparent that the luminescence was not fluorescence of the solid (such fluorescence is also observed on activated cathodes, but its spectrum is broad and diffuse). It appeared therefore that there were substantial quantities of Ba and Sr atoms in the interaction space. At least some of these were ionized, since some of the spectral lines observed were those of the positive, singly-charged ion.

In order to obtain further information about the atoms and ions in the anode-cathode space, "double-triode" tubes were built. In these tubes electrons from the central cathode bombarded both a nickel anode and an oxide-coated anode, which had been activated as if it were a cathode and which could be heated by a tungsten heater. A negative grid was placed in each half of the

---

[6] M. Benjamin and H. P. Rooksby, "Emission from Oxide-coated Cathodes," Phil. Mag. 15, 810 (1933).

tube, and the positive ion currents to these grids were measured. Both anodes were maintained at approximately the same potential, and the electron currents in the two parts of the tube were the same. The grid current in the nickel anode space was about 0.01 microampere when 10 milliamperes electron current was flowing; this corresponds to roughly $10^{-6}$ mm pressure of residual gas in the tube. The grid current in the other side of the tube was always many times this value, and when the oxide-coated anode was heated to 700–800°C, grid currents of 10–40 microamperes were observed. It was apparent, therefore, that atoms or ions were released from the bombarded surface, and that these formed a condensable gas phase which could not be observed outside the immediate neighborhood of the bombarded surface. Noise measurements were also made on these tubes; excess noise was observed, and it was substantially proportional to the ion current collected by the grid.

From these experiments it was possible to derive the following interpretation of the phenomena of magnetron noise: (1) Some of the electrons (of the order of 20 percent) emitted from the cathode return to it after gaining considerable energy from the r-f field. If the cathode is oxide-coated, these back-bombarding electrons remove atoms or ions of the cathode material. Any ions emitted quickly return to the cathode, but they alter the space charge in the region of the potential minimum and permit a momentary excess of electrons to be emitted. Atoms emitted have between a 1 percent and a 10 percent chance of being ionized before they strike the anode or other parts of the tube. Ions thus produced alter the space charge and may release other atoms or ions when they bombard the cathode. (2) Noise in excess of ordinary shot noise is produced by these processes. The noise occurring in space-charge limited electron currents when appreciable numbers of positive ions are present has been investigated by Thompson and North.[7] Most of the noise effects observed in the present tests of special triodes are consistent with their work. In general, however, the excess noise in magnetrons is greater than that obtained in any other vacuum tubes, and it would be necessary to assume pressures of the order of 0.1 to 0.01 mm of mercury in a magnetron in order to explain the observed excess noise by simple extrapolation of the work of Thompson and North. (3) The very great noise in magnetrons and the irregular dependence of noise upon time, current, and voltage, may be the result of cumulative or chain effects. The equivalent pressure of barium and strontium atoms near a hot magnetron cathode is of the order of $10^{-2}$ mm; this figure is derived from measurements on the "double-triode" tubes and is checked by the fact that at neon pressures less than $5\times10^{-3}$ mm the noise in an oxide-coated cathode magnetron was independent of pressure. At such a large pressure the contribution of positive ions to the space-charge distribution is considerable, and increasing the pressure by a factor of ten would cause oscillation to cease and a gas discharge at low voltage to occur. Under operating conditions the magnetron may therefore be on the verge of a Townsend breakdown. The ionization of a single atom may result in a burst of charge reaching the anode that is many times the electronic charge, and a correspondingly large increase in noise may occur. Among the processes which could participate in such a breakdown are ionization by electron impact, removal of atoms from the cathode by electrons, removal of atoms from the anode surface (which is soon contaminated with cathode material) by electrons, secondary electron emission at the cathode, and removal of atoms and electrons from the cathode surface by positive ion bombardment. It is not clear why such a breakdown does not occur, however, instead of merely threatening to occur.

Our interpretation is therefore that the dominating cause of the excess noise is associated with the presence of gas atoms in the interaction space. In a tungsten filament tube, these atoms are of residual gases and by sufficiently extensive pumping and gettering the excess noise may be substantially removed. In an oxide-coated cathode tube which is well out-gassed and gettered, the atoms are at least predominantly atoms removed from the cathode by electron back-bombardment.

The correlations reported in the preceding section give some support to these conclusions. The first correlation (noise *vs.* degree of back-bombardment) is evidently related to the de-

---

[7] B. J. Thompson and D. O. North, "Fluctuations Caused by Collision Ionization," RCA Review 5, 88 (1941).

pendence of the rate of removal of cathode material upon the number of bombarding electrons and the cathode temperature, both of which are higher in magnetrons with excessive back-bombardment. Experiments showed that the equivalent pressure of cathode material was an exponentially increasing function of the cathode temperature. Therefore, under conditions of severe back-bombardment, it would be expected that noise would be greater than with moderate bombardment. This is just what was observed.

The reduction in excess noise by reducing the sprayed length of the cathode is presumed to be related to the first correlation. In several tubes with short-spray-length cathodes, the cathodes operated at temperatures less than 700°C. Most full-length-sprayed cathodes operated at considerably higher temperatures. This is evidence of smaller back-bombardment when the electrons do not interact with electric fields near the ends of the anode.

The third correlation is still obscure. It may be that a very active thermionic cathode, which may have a surface layer of barium or strontium atoms, yields more atoms or ions when bombarded with electrons than a less active cathode.

A number of apparently unconnected relations are consistent with the gas-ionization hypothesis of the cause of excess noise: The decrease of noise with increasing I.F. frequency, the decrease of noise when the cathode heater power is removed, the negative results of various experiments with end-plate potentials and geometries, observations of the appearance of anode and cathode surfaces after many hours of operation, the occurrence of excess noise in magnetrons of quite different construction, the lack of dependence of noise phenomena upon cathode and anode diameters, and the irregular changes of noise with time, are all susceptible to explanation by this hypothesis.

There are three outstanding difficulties: (1) The physical process whereby atoms or ions are released from the cathode coating by bombarding electrons is not known, though several processes seem to be capable of producing this result. (2) The exact way in which ions and ionization of emitted atoms combine with the electron flow and the electric and magnetic fields to produce such large amounts of noise is not known. (3) The decrease of noise with increasing anode current does not seem to be explained.

Despite these difficulties, the hypothesis of the cause of the noise is felt to be reasonably well established. More detailed information is required about the mechanism of release of cathode material in order to advance our understanding of noise. Experiments with a mass spectrometer are planned to determine what molecules, atoms, or ions are released. The results of these experiments may make possible a quantitative theory of noise generation. Preliminary results with "double-triode" tubes indicate that at least some of the cathode material is removed as atoms and that the exponential increase of removed atoms (per unit electron current) has an exponent which corresponds to an "activation energy" of 40–50 kilocalories per mole. The heats of sublimation of barium and strontium are 44 and 38 kilocalories. This may be a coincidence, or it may be that electron bombardment produces small patches of barium and strontium, from which atoms readily volatilize at the cathode operating temperature.

## IV. PROGRESS IN THE REDUCTION OF EXCESS NOISE

The increase in $V_{a0}$ by reducing the sprayed length of the cathode has already been reported. This change and the improvement of the efficiency and reduction of back-bombardment resulting from further development of the frequency-modulated magnetrons[2] resulted in a large percentage of magnetrons which were not "noisy" in the ordinary operating region, near $V_a = 850$ volts.

The hypothesis of the cause of noise which was advanced in the preceding section suggested that the cathode be modified in order to reduce the excess noise and to raise $V_{a0}$ still further. No hope was entertained of eliminating the back-bombarding electrons, so a means was sought for preventing the removal of cathode material by such electrons. There is considerable evidence that electrons returning to the cathode do not strike at normal incidence, but at angles varying between 45° and tangential incidence. Furthermore, for a fixed magnetic field, the sense of rotation of all electrons about the cathode is the same. These facts permit the construction of a cathode such that returning electrons strike a nickel surface instead of striking the cathode

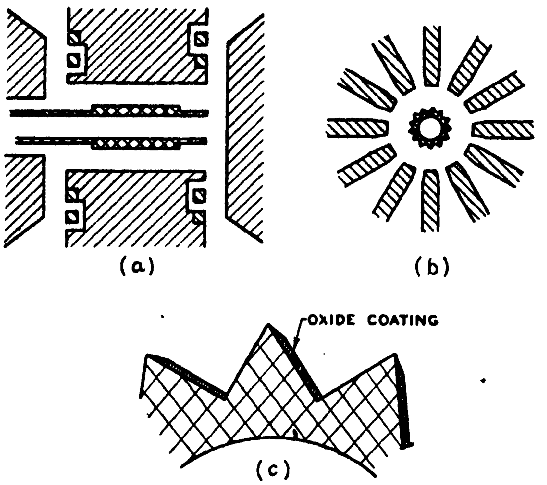


FIG. 4. Grooved cathode.

coating. One such cathode is shown in Fig. 4. The active length of the cathode is in the form of a 13-pointed star; only one side of each star point is coated with oxides. To accomplish this, the entire cathode was coated, and the spray was carefully scraped from areas where it was not wanted. This process was tedious, since the diameter of the cathode was only 0.070 inches, but was quite practical for experimental tubes.

Figure 5 shows the noise behavior of one of the four tubes constructed with the cathode of Fig. 4. By "correct" direction of the magnetic field we mean that direction which allows back-bombarding electrons to strike only the nickel surface (this is the clockwise sense of rotation in Fig. 4). It is evident that the noise is much less if the oxide-coating is sheltered from bombardment. This provides increased confidence in the hypothesis of the cause of noise which was advanced in the preceding section.

Three other tubes with cathodes as shown in Fig. 4 were built. One of these showed somewhat less noise for the "incorrect" direction of magnetic field than for the "correct" direction. This tube exhibited very high back-bombardment, more noise than other tubes of this type, and low efficiency. When it was dissected this tube was found to have a cathode which was not parallel to the axis of the anode. These facts make the information from this test of questionable value, but it is still not understood why this tube exhibited more noise when the electrons were rotating in the proper sense.

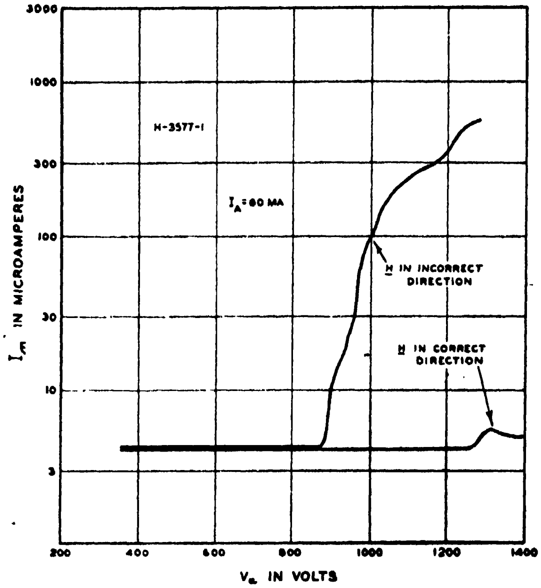


FIG. 5. Noise in magnetron with grooved cathode. When the magnetic field is in the "correct" direction, the oxide-coated surfaces are sheltered from back-bombarding electrons.

The remaining two tubes showed substantially the same behavior as the first (Fig. 5) and therefore confirmed the conclusions from that tube. The net result of all such experiments was that $V_{ao}$ is 200–300 volts larger when the magnetic field is applied in the "correct" sense.

An interesting possibility of cathodes of the general type of Fig. 4 is that they may permit substantially longer life than has previously been experienced with magnetrons. There is much evidence (some of it contained in the present work) that cathode life in a magnetron is limited more by the destruction of the cathode by back-bombardment than by the processes of cathode decay observed in ordinary diodes and negative grid tubes. If long life, not low excess noise, were the chief requirement, the oxide-coating could initially cover the entire cathode. Removal of part of the surface by back-bombardment would not cause failure, because at least part of the surface would be sheltered from back-bombarding electrons.

It is a privilege to acknowledge the cooperation of Messrs. G. R. Kilgore, C. I. Shulman, and Dr. J. Kurshan in this study, and the suggestions and encouragement of Dr. L. P. Smith and Dr. I. Wolff.

# A Semi-Polar Form of Fourier Series and Its Use in Crystal Structure Analysis

JOSEPH S. LUKESH
*Crystallographic Laboratory, Department of Geology, Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Received September 27, 1946)

A new form of Fourier series for crystal structure analysis is developed and a graphical method of summation described. The procedure is designed to permit simple and fairly rapid computation and recording of the density contributions of all planes at all points in the unit cell. Its particular advantage is in the case of centrosymmetrical projections where phases must be assigned experimentally. Adjustment of phases involves only re-addition of numbers rather than a complete new summation. The method also requires a minimum of equipment which is easily constructed at negligible expense.

THE distribution of electron density throughout the unit cell of a crystal can be determined by the summation of a Fourier series of the following general type

$$\rho(xyz)=\frac{1}{V}\sum_h\sum_k\sum_l F_{(hkl)}e^{-(2\pi i/N)(hx+ky+lz)} \qquad (1)$$

at each point, $xyz$, for all values of $F_{(hkl)}$. The coordinates, $xyz$, are given in integral numbers of $N$, the number of subdivisions of the axes. On expansion, Eq. (1) becomes

$$\rho(xyz)=\frac{1}{V}\left[\sum_h\sum_k\sum_l A_{(hkl)}\frac{\cos 2\pi}{N}(hx+ky+lz)\right.$$
$$\left.+\sum_h\sum_k\sum_l B_{(hkl)}\frac{\sin 2\pi}{N}(hx+ky+lz)\right]. \qquad (2)$$

In most crystal pattern syntheses, the series is restricted to a two-dimensional one whose summation gives a projection of the electron density. Considering only the reflections, $(hk0)$, for instance, the series is

$$\rho(xy)=\frac{1}{A}\left[\sum_h\sum_k A_{(hk0)}\frac{\cos 2\pi}{N}(hx+ky)\right.$$
$$\left.+\sum_h\sum_k B_{(hk0)}\frac{\sin 2\pi}{N}(hx+ky)\right]. \qquad (3)$$

Summation of Eq. (3) at all points, $xy$, gives a representation of the electron density projected on 001. Projections on 100 and 010 are obtained by considering only the $0kl$ and $h0l$ reflections, respectively.

Because of the tedious nature of the computations involved in even the two-dimensional series, a number of computational aids have been devised. All of these aids are characterized by the trigonometric separation of Eq. (3) into a group of one-dimensional series which can be summed easily by means of strips on which are printed the products of all possible magnitudes of $A$ and $B$ times all possible values of the cosine and sine. Summation is reduced to the addition of previously computed numbers.

Use of the computational aids reduces the labor of any one summation, but it may actually increase the total labor of a crystal structure determination. If one is dealing with a centrosymmetrical projection, the phases are all either 0 or $\pi$ and the equation to be summed is

$$\rho(xy)=\frac{1}{A}\sum_h\sum_k F_{(hk0)}\cos 2\pi(hx+ky). \qquad (4)$$

If the phase of a given reflection is 0, the sign of $F$ is positive; if it is $\pi$, the sign is negative. In most structure determinations, the phases are unknown and must be determined by trial and error. Given a trial structure, one may compute phases and these may then be used in the summation. In general, many of the signs will be found to have been incorrectly determined and the electron density will not be a faithful representation of the structure. It will then be necessary to change phases either experimentally or by making adjustments in the trial structure and re-computing. Each time phases are changed, a complete new summation must be made. When accurate parameters are being sought, a number

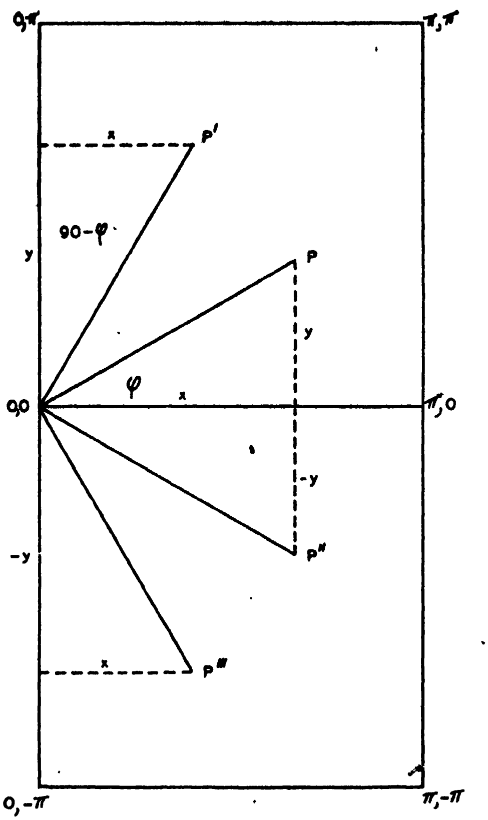


FIG. 1. The relation between rectilinear and semi-polar coördinates for points in a half-unit cell.

of approximations must usually be made before the final electron density map is obtained.

If, instead of summing Eq. (4) by use of strips, one employs a method in which the contributions of each plane at each point are determined individually and recorded, adjustment of phases will not require additional summations. To change signs, one need merely re-add columns of numbers. In the form of Eq. (4), the series requires a great deal of computation. However, if it is put into a new form, to be described in the following section, a graphical method of summation can be used which is both simple and reasonably rapid. In addition, the graphical method has an advantage in that the equipment required is a minimum, consisting of a chart, a scale, and a table of numbers.

### THEORY OF THE SEMI-POLAR SERIES

The point, $P$, in Fig. 1 is conventionally specified by the rectilinear coordinates $x$ and $y$. As an alternative, one may choose to locate $P$ by the coordinate $x$ and the angle, $\varphi$, between the $x$ axis and a line from the origin to $P$. Since

$$y = x\tan\varphi, \tag{5}$$

the rectilinear coordinate, $y$, may be replaced by $x\tan\varphi$. Equation (4) becomes

$$\rho(x, x\tan\varphi) = \frac{1}{A}\sum_h\sum_k F_{(hk0)} \times\cos\frac{2\pi}{N}(hx+kx\tan\varphi). \tag{6}$$

In a similar manner, the point, $P'$, in Fig. 1 may be specified by the coordinates $y\tan(90-\varphi)$ and $y$. In this case, the series is given by

$$\rho[y\tan(90-\varphi), y] = \frac{1}{A}\sum_h\sum_k F_{(hk0)} \times\cos\frac{2\pi}{N}[hy\tan(90-\varphi)+ky]. \tag{6'}$$

For all of the points along a line of constant $\varphi$ or $90-\varphi$, the tangent is a constant. Equations (6) and (6′) may, therefore, be written

$$\rho(x, c) = \frac{1}{A}\sum_h\sum_k F_{(hk0)} \cos\frac{2\pi}{N}x(h+ck), \tag{7}$$

$$\rho(c', y) = \frac{1}{A}\sum_h\sum_k F_{(hk0)} \cos\frac{2\pi}{N}y(c'h+k), \tag{7'}$$

where $c$ and $c'$ are the tangents of $\varphi$ and $90-\varphi$. For the points, $P''$ and $P'''$, the corresponding expressions are

$$\rho(x, -c) = \frac{1}{A}\sum_h\sum_k F_{(hk0)} \cos\frac{2\pi}{N}x(h-ck), \tag{7''}$$

$$\rho(-c', -y) = \frac{1}{A}\sum_h\sum_k F_{(hk0)} \times\cos\frac{2\pi}{N}(-y)(-c'h+k). \tag{7'''}$$

The expressions are the same, except that $c$, $c'$, and $y$ are negative. Since the cell is centrosymmetrical, only half need be considered.

The important feature of Eq. (7) and its related expressions is that for any given reflection, *hk*0, and any given slope, *c*, the entire expression in parentheses is a constant. Equation (7) may, then, be written

$$\rho(xc)=\frac{1}{A}\sum_h\sum_k F_{(hk0)}\cos\frac{2\pi}{N}xC. \qquad (8)$$

The density contribution of any plane at each point, *x*, along the line whose slope is *c* is found by multiplying $F$ by the cosine of $(2\pi C/N)\cdot 0$ (for $x=0$), $(2\pi C/N)\cdot 1$ (for $x=1$), $(2\pi C/N)\cdot 2$ (for $x=2$), etc. If one is interested in the density at any one point, *x*, *c*, Eq. (8) reduces to

$$\rho(xc)=\frac{1}{A}\Big[F_{(hk0)}\cos\frac{2\pi}{N}xC_{(hk0)} + F_{(h'k'0)}\cos\frac{2\pi}{N}xC_{(h'k'0)} + F_{(h''k''0)}\cos\frac{2\pi}{N}xC_{(h''k''0)}+\cdots\Big]. \qquad (9)$$

## A POLAR GRID

If the points on the rectilinear grid that is usually used in structure analysis are specified in the semi-polar form just described, there will be

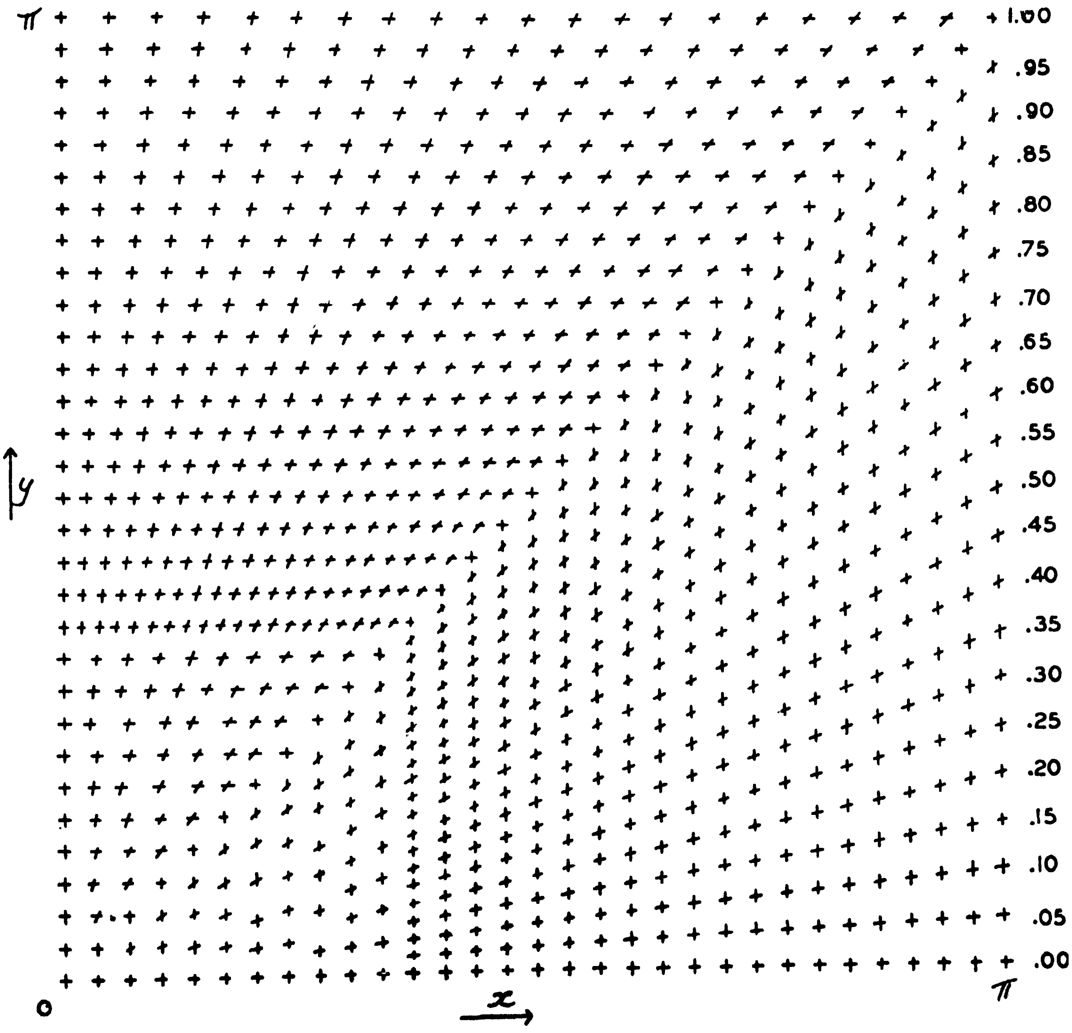


FIG. 2. The polar grid for one-quarter of the unit cell.

TABLE I. Values of $(2\pi/N)\cdot Cx$ for $C=4.80$.

| $x$ | $(2\pi/N)\cdot Cx$ |
|---|---|
| 0 | $(2\pi/N)\cdot 0.00$ |
| 1 | $(2\pi/N)\cdot 4.80$ |
| 2 | $(2\pi/N)\cdot 9.60$ |
| 3 | $(2\pi/N)\cdot 14.40$ |
| 4 | $(2\pi/N)\cdot 19.20$ |
| 5 | $(2\pi/N)\cdot 24.00$ |
| 6 | $(2\pi/N)\cdot 28.80$ |
| 7 | $(2\pi/N)\cdot 33.60=(2\pi/N)\cdot 26.40$ |
| 8 | $(2\pi/N)\cdot 38.40=(2\pi/N)\cdot 21.60$ |

TABLE II. Product of $F$ times $\cos(2\pi/N)\cdot Cx$.

| $x$ | $Cx$ | $F\cos(2\pi/N)\cdot Cx$ |
|---|---|---|
| 0 | 0.00 | 100 |
| 1 | 4.80 | 89 |
| 2 | 9.60 | 53 |
| 3 | 14.40 | 6 |
| 4 | 19.20 | −43 |
| 5 | 24.00 | −80 |
| 6 | 28.80 | −99 |
| 7 | 26.40 | −95 |
| 8 | 21.60 | −64 |

a large number of constants, $c$, corresponding to the tangents of the various angles, and, in general, their numerical values will not be simple numbers. Because of this, computation of the constant, $C$, will not be easy. Furthermore, there will be only a few points along each line, except in special cases. The value of having a "constant" will be reduced. Both of these difficulties can be eliminated if one abandons the conventional grid in favor of one defined by the intersections of the $x$ and $y$ coordinates with radial lines of simple slopes. For each radial line, there will be $N/2$ points, assuming the summation is made from the origin to $\pi$.

A grid suitable for most purposes is formed by allowing $N$ to equal 60. There are, then, thirty divisions along each axis from 0 to $\pi$. To approximate the distribution of points formed by such a rectilinear grid, a series of radial lines whose slopes are .05, .10, .15, .20, etc. has been chosen. Intersections of these with the $x$ coordinates locate the points of the grid for the first segment, in Fig. 1, of the cell. For the second segment, a symmetrical array is obtained by using the intersections of the $y$ coordinates and radial lines whose slopes, $c'$, have the same values. The grid thus defined for the first two segments of the cell is shown in Fig. 2. The principal disadvantage, outside of the unconventional array, lies in the fact that there are only twenty points along each of the axes, $x=\pi$ and $y=\pi$. To offset this, points half-way between those shown may be handled by interpolation. The point density near the origin has been reduced by eliminating some of the points, those from the origin to $x$ and $y$ equal four being, actually, coincident with points of the conventional grid.

It might be well at this time to discuss, briefly, the mathematics of the semi-polar coordinate system just developed. Heretofore, the tacit assumption has been made that the unit cell is square so that a point at $x=10$ and $y=5$, for instance, actually lies along a radius which makes an angle with the $x$ axis whose tangent is .50. It may not be immediately obvious that the same relation holds for a rectangular or oblique cell. Actually, the expression, $\tan\varphi=y/x$, is merely a convenient mathematical shorthand for referring to the ratio of units of distance along the $y$ axis to that along the $x$ axis. No matter what the absolute lengths of the two units may be or the angle between the axes, the statement, $\tan\varphi=.50$, simply means that the point is twice as many $x$ units from the origin as $y$ units. This holds, of course, only when the axes are divided into the same number of parts.

## GRAPHICAL SUMMATION OF THE SEMI-POLAR SERIES

As was mentioned in an earlier section, for a given plane, $hk0$, and a radial line of slope, $c$, the term, $(h+ck)$, in Eq. (7) is a constant for all points along the line, and the series may be expressed in the form of Eq. (8). The density contribution of $hk0$ at each point of the grid along the radial line is computed by multiplying the amplitude, $F$, by the cosines of a series of angles which are multiples of $2\pi C/N$. The discrete values of the cosine so obtained may be computed, once for all, and tabulated. Since they depend only on the geometry of the plane and are independent of the structure, such a table could be used for any structure determination. However, use of a table of cosines would require that multiplication be performed for each plane at each point, which is an extremely tedious task. Inasmuch as Eq. (8) is a periodic cosine function, the product can be determined readily by graphical means. Given a cosine curve and a vertical scale appropriate to the magnitude of $F$, it is only necessary to know what points along the axis of the curve correspond to multiples of $2\pi C/N$. These, of course, can be determined in advance and tabulated.

As an example, the density contributions of the plane, 420, along the radial line whose slope is .40 will be determined. The amplitude will be assumed to be +100. The constant, $C$, in Eq. (8) is given by

$$\begin{aligned} C &= h + ck, \\ &= 4 + .40 \cdot 2, \\ &= 4.80. \end{aligned}$$

The values of the product, $2\pi C/N$ times $x$ for the first eight points along the radial line are listed in Table I. Because the cosine function is sym-

TABLE III.* Part of the tabulation of $C$ times $x$ as a function of $C$ and $x$.

| $C$ \ $x$=0 | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4.00 | 0.00 | 4.00 | 8.00 | 12.00 | 16.00 | 20.00 | 24.00 | 28.00 | 28.00 | 24.00 | 20.00 | 16.00 | 12.00 | 8.00 | 4.00 |
| 4.05 | 0.00 | 4.05 | 8.10 | 12.15 | 16.20 | 20.25 | 24.30 | 28.35 | 27.60 | 23.55 | 19.50 | 15.45 | 11.40 | 7.35 | 3.30 |
| 4.10 | 0.00 | 4.10 | 8.20 | 12.30 | 16.40 | 20.50 | 24.60 | 28.70 | 27.20 | 23.10 | 19.00 | 14.90 | 10.80 | 6.70 | 2.60 |
| 4.15 | 0.00 | 4.15 | 8.30 | 12.45 | 16.60 | 20.75 | 24.90 | 29.05 | 26.80 | 22.65 | 18.50 | 14.35 | 10.20 | 6.05 | 1.90 |
| 4.20 | 0.00 | 4.20 | 8.40 | 12.60 | 16.80 | 21.00 | 25.20 | 29.40 | 26.40 | 22.20 | 18.00 | 13.80 | 9.60 | 5.40 | 1.20 |
| 4.25 | 0.00 | 4.25 | 8.50 | 12.75 | 17.00 | 21.25 | 25.50 | 29.75 | 26.00 | 21.75 | 17.50 | 13.25 | 9.00 | 4.75 | 0.50 |
| 4.30 | 0.00 | 4.30 | 8.60 | 12.90 | 17.20 | 21.50 | 25.80 | 29.90 | 25.60 | 21.30 | 17.00 | 12.70 | 8.40 | 4.10 | 0.20 |
| 4.35 | 0.00 | 4.35 | 8.70 | 13.05 | 17.40 | 21.75 | 26.10 | 29.55 | 25.20 | 20.85 | 16.50 | 12.15 | 7.80 | 3.45 | 0.90 |
| 4.40 | 0.00 | 4.40 | 8.80 | 13.20 | 17.60 | 22.00 | 26.40 | 29.20 | 24.80 | 20.40 | 16.00 | 11.60 | 7.20 | 2.80 | 1.60 |
| 4.45 | 0.00 | 4.45 | 8.90 | 13.35 | 17.80 | 22.25 | 26.70 | 28.85 | 24.40 | 19.95 | 15.50 | 11.05 | 6.60 | 2.15 | 2.30 |
| 4.50 | 0.00 | 4.50 | 9.00 | 13.50 | 18.00 | 22.50 | 27.00 | 28.50 | 24.00 | 19.50 | 15.00 | 10.50 | 6.00 | 1.50 | 3.00 |
| 4.55 | 0.00 | 4.55 | 9.10 | 13.65 | 18.20 | 22.75 | 27.30 | 28.15 | 23.60 | 19.05 | 14.50 | 9.95 | 5.40 | 0.85 | 3.70 |
| 4.60 | 0.00 | 4.60 | 9.20 | 13.80 | 18.40 | 23.00 | 27.60 | 27.80 | 23.20 | 18.60 | 14.00 | 9.40 | 4.80 | 0.20 | 4.40 |
| 4.85 | 0.00 | 4.65 | 9.30 | 13.95 | 18.60 | 23.25 | 27.90 | 27.45 | 22.80 | 18.15 | 13.50 | 8.85 | 4.20 | 0.45 | 5.10 |
| 4.70 | 0.00 | 4.70 | 9.40 | 14.10 | 18.80 | 23.50 | 28.20 | 27.10 | 22.40 | 17.70 | 13.00 | 8.30 | 3.60 | 1.10 | 5.80 |
| 4.75 | 0.00 | 4.75 | 9.50 | 14.25 | 19.00 | 23.75 | 28.50 | 26.75 | 22.00 | 17.25 | 12.50 | 7.75 | 3.00 | 1.75 | 6.50 |
| 4.80 | 0.00 | 4.80 | 9.60 | 14.40. | 19.20 | 24.00 | 28.80 | 26.40 | 21.60 | 16.80 | 12.00 | 7.20 | 2.40 | 2.40 | 7.20 |
| 4.85 | 0.00 | 4.85 | 9.70 | 14.55 | 19.40 | 24.25 | 29.10 | 26.05 | 21.20 | 16.35 | 11.50 | 6.65 | 1.80 | 3.05 | 7.90 |
| 4.90 | 0.00 | 4.90 | 9.80 | 14.70 | 19.60 | 24.50 | 29.40 | 25.70 | 20.80 | 15.90 | 11.00 | 6.10 | 1.20 | 3.70 | 8.60 |
| 4.95 | 0.00 | 4.95 | 9.90 | 14.85 | 19.80 | 24.75 | 29.70 | 25.35 | 20.40 | 15.45 | 10.50 | 5.55 | 0.60 | 4.35 | 9.30 |
| 5.00 | 0.00 | 5.00 | 10.00 | 15.00 | 20.00 | 25.00 | 30.00 | 25.00 | 20.00 | 15.00 | 10.00 | 5.00 | 0.00 | 5.00 | 10.00 |

TABLE III. (*Continued*)

| 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.00 | 4.00 | 8.00 | 12.00 | 16.00 | 20.00 | 24.00 | 28.00 | 28.00 | 24.00 | 20.00 | 16.00 | 12.00 | 8.00 | 4.00 | 0.00 |
| 0.75 | 4.80 | 8.85 | 12.90 | 16.95 | 21.00 | 25.05 | 29.10 | 26.85 | 22.80 | 18.75 | 14.70 | 10.65 | 6.60 | 2.55 | 1.50 |
| 1.50 | 5.60 | 9.70 | 13.80 | 17.90 | 22.00 | 26.10 | 29.80 | 25.70 | 21.60 | 17.50 | 13.40 | 9.30 | 5.20 | 1.10 | 3.00 |
| 2.25 | 6.40 | 10.55 | 14.70 | 18.85 | 23.00 | 27.15 | 28.70 | 24.55 | 20.40 | 16.25 | 12.10 | 7.95 | 3.80 | 0.35 | 4.50 |
| 3.00 | 7.20 | 11.40 | 15.60 | 19.80 | 24.00 | 28.20 | 27.60 | 23.40 | 19.20 | 15.00 | 10.80 | 6.60 | 2.40 | 1.80 | 6.00 |
| 3.75 | 8.00 | 12.25 | 16.50 | 20.75 | 25.00 | 29.25 | 26.50 | 22.25 | 18.00 | 13.75 | 9.50 | 5.25 | 1.00 | 3.25 | 7.50 |
| 4.50 | 8.80 | 13.10 | 17.40 | 21.70 | 26.00 | 29.70 | 25.40 | 21.10 | 16.80 | 12.50 | 8.20 | 3.90 | 0.40 | 4.70 | 9.00 |
| 5.25 | 9.60 | 13.95 | 18.30 | 22.65 | 27.00 | 28.65 | 24.30 | 19.95 | 15.60 | 11.25 | 6.90 | 2.55 | 1.80 | 6.15 | 10.50 |
| 6.00 | 10.40 | 14.80 | 19.20 | 23.60 | 28.00 | 27.60 | 23.20 | 18.80 | 14.40 | 10.00 | 5.60 | 1.20 | 3.20 | 7.60 | 12.00 |
| 6.75 | 11.20 | 15.65 | 20.10 | 24.55 | 29.00 | 26.55 | 22.10 | 17.65 | 13.20 | 8.75 | 4.30 | 0.15 | 4.60 | 9.05 | 13.50 |
| 7.50 | 12.00 | 16.50 | 21.00 | 25.50 | 30.00 | 25.50 | 21.00 | 16.50 | 12.00 | 7.50 | 3.00 | 1.50 | 6.00 | 10.50 | 15.00 |
| 8.25 | 12.80 | 17.35 | 21.90 | 26.45 | 29.00 | 24.45 | 19.90 | 15.35 | 10.80 | 6.25 | 1.70 | 2.85 | 7.40 | 11.95 | 16.50 |
| 9.00 | 13.60 | 18.20 | 22.80 | 27.40 | 28.00 | 23.40 | 18.80 | 14.20 | 9.60 | 5.00 | 0.40 | 4.20 | 8.80 | 13.40 | 18.00 |
| 9.75 | 14.40 | 19.05 | 23.70 | 28.35 | 27.00 | 22.35 | 17.70 | 13.05 | 8.40 | 3.75 | 0.90 | 5.55 | 10.20 | 14.85 | 19.50 |
| 10.50 | 15.20 | 19.90 | 24.60 | 29.30 | 26.00 | 21.30 | 16.60 | 11.90 | 7.20 | 2.50 | 2.20 | 6.90 | 11.60 | 16.30 | 21.00 |
| 11.25 | 16.00 | 20.75 | 25.50 | 29.75 | 25.00 | 20.25 | 15.50 | 10.75 | 6.00 | 1.25 | 3.50 | 8.25 | 13.00 | 17.75 | 22.50 |
| 12.00 | 16.80 | 21.60 | 26.40 | 28.80 | 24.00 | 19.20 | 14.40 | 9.60 | 4.80 | 0.00 | 4.80 | 9.60 | 14.40 | 19.20 | 24.00 |
| 12.75 | 17.60 | 22.45 | 27.30 | 27.85 | 23.00 | 18.15 | 13.30 | 8.45 | 3.60 | 1.25 | 6.10 | 10.95 | 15.80 | 20.65 | 25.50 |
| 13.50 | 18.40 | 23.30 | 28.20 | 26.90 | 22.00 | 17.10 | 12.20 | 7.30 | 2.40 | 2.50 | 7.40 | 12.30 | 17.20 | 22.10 | 27.00 |
| 14.25 | 19.20 | 24.15 | 29.10 | 25.95 | 21.00 | 16.05 | 11.10 | 6.15 | 1.20 | 3.75 | 8.70 | 13.65 | 18.60 | 23.55 | 28.50 |
| 15.00 | 20.00 | 25.00 | 30.00 | 25.00 | 20.00 | 15.00 | 10.00 | 5.00 | 0.00 | 5.00 | 10.00 | 15.00 | 20.00 | 25.00 | 30.00 |

---

* *Explanation of Table III.* This table contains that portion of the tabulation of the product of the constant, $C$, times the coordinate, $x$, for $C$ between 4.00 and 5.00, inclusive. The values listed are obtained by multiplying $C$ by $x$, any product larger than 30 being reduced to an equivalent value less than 30. The remainder of the tabulation, covering the rest of the [illegible] The values listed in the table are the coordinates along the axis of a cosine curve, graduated in terms of $2\pi/60$, at which a scale should be set in order to determine graphically the product of the amplitude times the appropriate cosine.

metrical about 0 and $\pi$, one-half a cycle of the curve is sufficient if all values of the product greater than $\pi$ are recomputed to a symmetrical value less than $\pi$. In the present case, the product is equal to $\pi$ when $C$ times $x$ is equal to 30.

The products listed in Table I are the angles by whose cosines the amplitude must be multiplied to determine the density contribution of the plane, 420, at each of the first eight points of the grid along the radial line whose slope is .40. If the cosine curve is calibrated in terms of $2\pi/N$, it is only necessary to set the scale at the points 4.80, 9.60, 14.40, etc. and to read off the product. The values so obtained are listed in Table II.

In order to perform the computations for all planes at all points in the cell, it is necessary to tabulate all possible solutions of $x$ times $(h+ck)$ and the related functions. This might at first seem to require an extensive table since the values of $h$ and $k$ may become quite large in large cells. However, the symmetry of the cosine function reduces the number of independent solutions to a finite number. The highest independent value of the product of $C$ times $x$, and hence of $C$ itself, is that which makes the product $(2\pi C/N)\cdot x$ equal to $\pi$. In the present case, the value is 30. Any value of $C$ greater than thirty is symmetrical with one less than thirty. This is evident from $x=7$ and $x=8$ in Table I. Furthermore, since the cosine function is symmetrical about 0, negative value of $C$ are the same as positive, and the sign of $C$ may be disregarded. With $N$ equal to 60, there are only 600 possible values of $C$, and the products of $C$ times $x$ can be tabulated conveniently. Such a table can be used for all structure determinations. A portion of the table is shown in Table III.

In many cases, it is sufficient to base a grid on $N$ equal to 30. Radial lines of slopes .1, .2, .3, etc. may be used. The tabulation of $C$ times $x$ requires only 150 values of $C$ against 15 values of $x$.

## SUMMATION PROCEDURE

In determining the electron density by the method just developed, it is convenient to compute the density contributions of each plane at each point along a given radial line at the same time, as was done in the example. Worksheets can be prepared for each radial line on which are columns for the indices, the constants for each reflection, the $F$ values, and for the density contributions at each $x$ coordinate. The contributions of each plane, in turn, can then be easily determined by setting the proper scale at the points along the cosine curve indicated in the tabulation of $C$ times $x$. When all planes have been handled, the columns of density contributions are totaled, and the totals plotted on the polar grid. If the resulting electron density map indicates that adjustments must be made in phases, it is only necessary to re-add the columns.

## SUMMARY

A semi-polar form of Fourier series has been developed and a graphical method of summation described. Although the procedure is somewhat more tedious than the conventional strip methods of Fourier series summation, it may be found useful, particularly when phases must be assigned experimentally in centrosymmetrical projections. An added advantage is the simplicity of the equipment, which can be constructed at a negligible cost.

# The Insulation of High Voltages in Vacuum

JOHN G. TRUMP
*Department of Electrical Engineering*

AND

R. J. VAN DE GRAAFF
*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Received December 23, 1946)

Breakdown studies have been made between electrodes in high vacuum at constant voltages from 50 to 700 kv. These further demonstrate the inadequacy of the field emission theory to account generally for high voltage breakdown in vacuum. Experiments are described which investigate some of the "total voltage" breakdown mechanisms, including positive-ion emission by electron impact, electron emission by positive-ion impact and by photons. In the d.c. case these processes contribute to a steady interchange of charged particles between cathode and anode which increases with voltage until breakdown ensues. At higher breakdown voltages the cathode gradient has diminished far below the value for field emission. Measurements of electron emission by electrons with energies up to 300 kv for tungsten, steel, aluminum, and graphite are reported. The possibilities of predicting and of improving the insulating strength of electrode gaps in high vacuum by the study of the coefficients of the electrode materials are discussed.

## INTRODUCTION

THE use of vacuum for the insulation of high voltages has become of increasing importance in modern science. In physics, vacuum insulation is essential for the acceleration of charged particles to high energies for nuclear studies, and generally for the production of high energy radiations. In engineering, the primary application of vacuum insulation has been in electronic tubes, including high vacuum rectifiers and high voltage x-ray tubes. The present applications of vacuum to low power switches and low loss high frequency condensers merely suggest the possible future importance which vacuum insulation may have in electrical power engineering.

During the last decade, the technique of producing and maintaining vacuum in large volumes has developed to the extent that vacuum may now be regarded as an available voltage-insulating medium. Nevertheless, although solid, liquid, and gaseous dielectrics have been intensively studied for many years, relatively little effort has been directed toward understanding and developing the unique electrical insulating properties of vacuum. After interruption by the war, the study of vacuum insulation, particularly at high voltages, is being resumed in the Department of Electrical Engineering at M.I.T. Some of the information obtained in these prior studies, as well as comments on our present picture of the mechanism of voltage breakdown in vacuum, is reported herein.

Vacuum insulation may be said to exist when the breakdown strength between metallic electrodes is independent of the pressure of the intervening gas. In an insulating vacuum the mean free path of electrons and positive ions moving in the residual gas must, therefore, be large compared with the inter-electrode gap in order that the contribution of gaseous ionization to the breakdown mechanism be negligible. While this condition may be achieved at atmospheric pressure for gaps of a small fraction of the wave-length of light, for the more ordinary gaps of physics and engineering a residual gas pressure of $10^{-4}$ mm of Hg or less is generally adequate. Further reduction in pressure does not increase the insulating strength of the gap except insofar as residual gas pressure may affect the absorbed gas and molecular layers of contaminating material which are in equilibrium on the electrode surfaces.

Vacuum insulation is more susceptible to quantitative analysis than any other insulating medium. Being free from the obscurities and limitations of intervening materials, the situation between metallic electrodes immersed in high

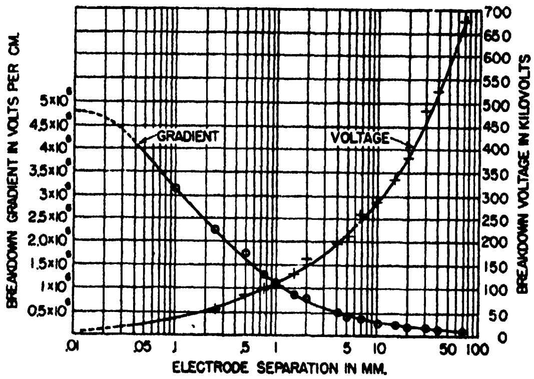


FIG. 1. Breakdown voltages and gradients between 1″ stainless steel ball and 2″ steel disk in vacuum.

vacuum may be more precisely described in terms of the electric field configuration, the number and energy of the charged particles, and other factors which influence the current flow and breakdown strength. Compared to a gas gap at high pressure, the phenomena in a vacuum gap are fewer and more direct. All of the ionization and recombination processes which occur in the gas are eliminated along with the field-distorting influence of space charge; only the electrodes remain effective in contributing or removing charged particles from the inter-electrode gap.

## FIELD EMISSION BREAKDOWN AT LOW VOLTAGES

Much work has been done on vacuum insulation at low d.c. voltages and high field strengths with the object of studying the phenomena of high field emission. These experiments in general were made using well-outgassed electrodes in a highly evacuated chamber, the electrode arrangement consisting variously of concentric cylinders with the inner cylinder a filament easily outgassed by heating, of a point directed at a plane, of two fine crossed wires, or of sphere gaps. These experiments indicated that gradients at the cathode surface of several times $10^6$ volts per centimeter could be insulated before breakdown ensued, and that gradients at least ten times higher than this value could be insulated on the anode surface. Hayden[1] reports a maximum cathode gradient of $1.3\times10^6$ volts per centimeter between outgassed molybdenum spheres. Piersol[2] observed a cathode gradient of $5.4\times10^6$ volts per centimeter between molybdenum plates outgassed at 1400°C. In the work described in the present paper it was found that at voltages less than 50 kilovolts it is quite easy to support gradients above $10^6$ volts per centimeter on steel cathode surfaces after conditioning the surface by permitting low energy flashover.

In the voltage range below 20 kilovolts detectable high field emission from cathode surfaces in vacuum begins at a fairly definite gradient which depends on the electrode material and its surface condition and increases exponentially with increasing gradient until flashover. This effect was first quantitatively analyzed by Schottky,[3] who modified the Richardson thermionic emission equation by reducing the true work function of the metal to account for the attractive influence of the electric field on the escaping electrons. The later experiments of Millikan[4] and others indicated that, unlike the prediction of the Schottky equation, high field emission is substantially independent of the cathode temperature, at least up to about 1200°C, beyond which it is overshadowed by the thermionic emission. This non-dependence has been more adequately explained by the newer Fowler-Nordheim relation.[5]

Both the Schottky and Fowler-Nordheim relations assume the existence of a smooth cathode surface. Actually, high field emission, like thermionic emission,[6] takes place preferentially from the microscopic projections, crystal edges, and spots of impurities which usually exist on surfaces. This consideration of surface roughness and contamination explains to a considerable extent the discrepancy between the experimentally observed cathode gradients and the higher values predicted by theory. Since field emission varies exponentially with gradient, it is evident that an intensification as small as 10

---

[1] Hayden, J. A. I. E. E. **41**, 854 (1922).

[2] Piersol, Rep. Brit. Assoc. Advancement Sci. 359 (1924).

[3] Schottky, *Jahrbuch der Radioactivitat und Elektrotechnik* (1915), Vol. 12, pp. 200–205.

[4] Millikan, Eyring, and McKeown, Phys. Rev. **31**, 900 (1928).

[5] Fowler and Nordheim, Roy. Soc. Proc. **A119**, 173–181 and **A121**, 626–639 (1928).

[6] R. P. Johnson and W. Shockley, Phys. Rev. **49**, 433 (1936).

percent because of surface projections would cause most of the field-emission current to flow from such points. The adverse influence on the breakdown gradient of such factors as contaminating surface films, rough electrode surfaces, and low work-function electrodes has become generally appreciated.

It is now well established that voltage breakdown between metal electrodes in high vacuum at high gradients, but relatively low voltages is initiated by high field emission. This emission increases exponentially with gradient and results in local temperature increases and enhanced secondary effects which lead to instability. For many years it has been generally believed that this same high field mechanism accounted for breakdown at high voltages. That this cannot be the case is demonstrated by the data of Fig. 1, in which breakdown voltage and gradient are shown as a function of electrode separation for constant voltages from 20 to 650 kilovolts. It is seen that, while fairly constant gradients of several million volts per centimeter are required to produce breakdown at voltages less than 50 kilovolts, the cathode gradient required at voltages higher than this decreases rapidly. Thus, whereas a gradient of 3000 kilovolts per centimeter is required when the total voltage between the electrodes is 20 kilovolts, a gradient of less than 100 kilovolts per centimeter is the maximum that can be supported at 650 kilovolts. This gradient is evidently inadequate of itself to insure breakdown through the mechanism of high field emission.

## HIGH VOLTAGE BREAKDOWN

The explanation of high voltage breakdown suggested some years ago by one of us[7] pictures an interchange of charged particles and photons between cathode and anode. Thus, an electron starting from the cathode and accelerated by the voltage between the electrodes, impinges on the anode with the resultant possible emission of both a positive ion and photons. Such positive ions and some of the photons would return to the cathode and cause further electron emission. When the conditions are such that this interchange becomes cumulative, breakdown results. To represent this mechanism quantitatively, let coefficient $A$ represent the average number of positive ions produced by one electron, and coefficient $B$ the average number of secondary electrons produced by one of these positive ions; similarly, let $C$ represent the photons produced by one electron, and $D$ the secondary electrons produced by an average photon. Then, when the condition

$$AB+CD>1$$

is realized, the interchange becomes cumulative and breakdown ensues. It is evident that these coefficients will be functions of the accelerating voltage between the electrodes and the gradient at the electrode surfaces, as well as of the electrode material and the surface condition. Some preliminary experiments to measure these coefficients were undertaken before the war with the collaboration of J. R. Maull, and work in this direction is now being resumed.

The experimental arrangement for determining coefficient $A$ is shown in Fig. 2. Electrons starting from the filament are accelerated by the voltage between electrodes, and impinge on the upper target, where they are measured. Because of the uniform field, any positive ions produced by these primary electrons return along the same path and are measured by collection in the

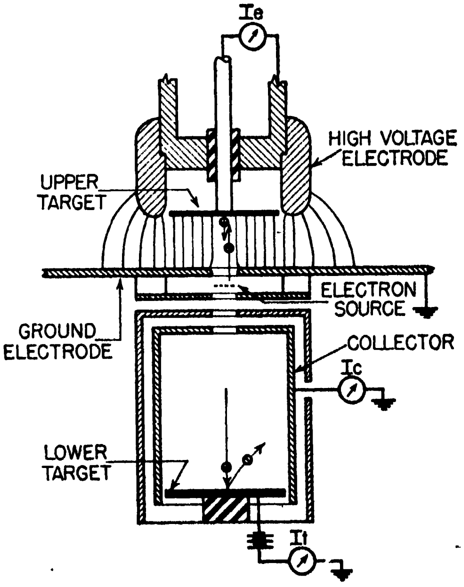


FIG. 2. Electrode arrangement for measuring "breakdown coefficients."

---

[7] Van Atta, Van de Graaff, and Barton, Phys. Rev. 43, 158 (1933).

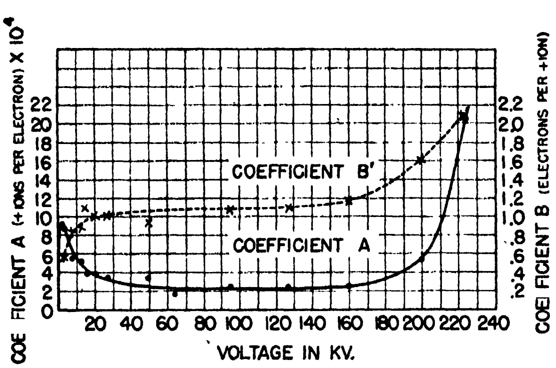


FIG. 3. Breakdown coefficients for steel electrodes with 2-cm gap.

Faraday cage. The ratio of the positive-ion current to the electron current is the coefficient $A$, and can thus be determined over a range of voltages and gradients and different electrode materials. It is noteworthy that the emission of positive ions due to electron impact is assisted in this experiment by the electric field at the anode surface. This is exactly the condition which exists in an actual gap in high vacuum.

Figure 3 shows the coefficient $A$, the ratio of secondary positive ions to electrons, for a parallel-plane steel gap as a function of the gap voltage. This curve was quite reproducible after the steel target had been outgassed by electron bombardment. The low voltage peak was found to be caused by the ionization of the residual gas and varied closely with its pressure. As would be expected, this residual gas ionization reduces with increasing gap voltage so that at the higher voltages the returning positive ions are owing primarily to secondary ion emission from the anode. The small proportion of residual gas ions to ions from the anode itself accounts for the small dependence on pressure of breakdown voltages in high vacuum.

That a definite positive-ion current could be measured from the electron bombarded target is evidence that this mechanism contributes to breakdown in vacuum. The small value of coefficient $A$, in this case 2 to $20\times10^{-4}$ in the energy range below breakdown, indicates the cathode coefficient $B$ must be large if this particle mechanism is to be important. The sharp increase in coefficient $A$ near the breakdown voltage is to be noted. The existence of secondary positive-ion emission is also supported by two observations of Anderson in the course of a doctorate research in this program.[8] With constant applied potential and a substantially uniform field, Anderson found deposition of anode (copper) material on the cathode after some minutes of operation near the breakdown value. He also showed a definite, though small, cathode temperature rise under such conditions—this temperature rise being a fraction of one percent of the rise of this same electrode when the polarities were reversed.

It should be expected that the coefficient $A$ will in general differ from one metal to another and show considerable dependence on surface condition.

The same arrangement shown in Fig. 2 may be used to give an idea of the coefficient $B$, the average number of electrons produced at the cathode by the impact of one high energy positive ion. The returning positive ions, having been accelerated by the full gap voltage, are permitted to impinge on a second target within the collector cage. The resultant secondary electrons are collected as current $I_c$ in this experiment. Figure 3 shows that under these conditions the secondary electrons increase slowly with the energy of the impacting positive ions, the ratio being a small number at energies below 220 kv. This result checks generally with the measurements of Allen,[9] Linford,[10] Hill *et al*,[11] in which secondary electron emission under high energy positive-ion bombardment and low cathode field were found to be small (2 to 20) and usually increased with voltage.

Since, in an actual gap, the cathode is subjected to a high electric gradient which would have the effect of assisting in the escape of electrons, the cathode coefficient shown in Fig. 3 is labeled $B'$ and must be regarded as low, perhaps low by the large factor required to make the particle mechanism of importance in vacuum breakdown. It is evident that a heavy positive ion having an energy of the order of a hundred kilovolts will produce brief, but extremely intense, local heating at its point of impact on the cathode. The effect of a high cathode gradient

---

[8] Anderson, Elec. Eng. 54, 1315 (1935).

[9] J. S. Allen, Phys. Rev. 55, 236 (1939).

[10] L. H. Linford, Phys. Rev. 47, 279 (1935).

[11] Hill, Buechner, Clark, and Fisk, Phys. Rev. 55, 463 (1939).

would be to reduce the space charge limitation to electron emission during this short period of intense local heat. Thus under the conditions of high cathode gradient a relatively large number of electrons might escape because of the impact of a single positive ion. Experiments are now planned to determine coefficient $B$ under the complete conditions of voltage and gradient.

Unless retarded by the external circuit, the collapse of the insulating strength of a vacuum gap is characterized by unusually rapid development greatly exceeding that of gaseous or solid insulator gaps. Evidently the conditions for instability are sharply defined and the subsequent discharge is relatively unimpeded. This is to be expected because of the short transit times involved in the energy interchange between electrodes and the absence of space-charge phenomena in the early phases. At an intermediate stage, after the condition for instability has been reached, self-focusing processes probably concentrate the discharge into a thin-cored channel and further enhance the emission of charged particles from the surfaces through thermionic emission. It has been observed that higher voltages can be insulated in the case of somewhat ununiform fields than when the interchange of particles can proceed undeviating along a straight line path between electrodes. The problem of vacuum insulation must be met primarily by reducing the emission characteristics of the electrode surfaces so as to prevent the onset of instability until well beyond the required conditions of voltage and gradient.

### EMISSION OF ELECTRONS BY HIGH ENERGY ELECTRON IMPACT

Primarily by replacing the upper target of Fig. 2 with a tungsten filament, the experimental set-up was adapted for measuring the number of secondary electrons emitted due to impact of high energy primary electrons, as a function of voltage and electrode material. Figure 4 shows the ratio of secondary electrons to primary electrons from 30 kv to 340 kv for tungsten, steel, aluminum, and graphite. It is seen that this ratio diminishes with increasing primary energy and that the ratio is a function of the atomic number of the electrode. A study of the energy of these secondary electrons showed a low energy group with energies less than 20 volts. Most of the remainder of the secondaries had energies in excess of 800 volts, some extending in the limit up to the energy of the primaries. As can be seen from the ascending curves of Fig. 4, the proportion of these high energy secondaries increased with voltage, becoming fairly constant above 200 kv. These more energetic electrons result from elastic nuclear scattering and from electron-electron collisions.

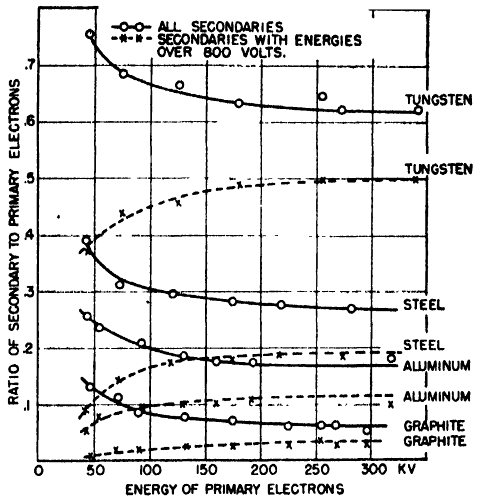


FIG. 4. Secondary emission of electrons by electrons with energies up to 340 kv.

To the picture of vacuum breakdown this information adds the concept that electrons from the cathode striking the anode cause repeated opportunities—though each time with diminished energy—to bring about the emission of positive ions and photons. The reduction of this secondary emission coefficient at the higher voltages illustrates how some breakdown factors may actually become diminished at extreme voltage values.

Experiments have been devised, but not yet carried out, for measuring the contribution of the photoelectric mechanism to high voltage breakdown in vacuum. It is well known that the efficiency of x-ray production increases markedly with voltage, though with an increasing tendency for the radiation to continue in the forward direction. Only backward-scattered radiation would be effective in releasing more electrons from the cathode. Although a high electric field at the cathode would assist the photoelectric emission, the probability of this mechanism being important appears to be small.

The several processes which contribute to the interchange of charged particles are seen to depend in an important way on the total available voltage between the vacuum-immersed electrodes. The breakdown voltage curve of Fig. 1 clearly shows that, at the higher voltage values across a given gap, it is the voltage and not primarily the gradient that affects the insulation performance. For these reasons the term "total voltage mechanism" has been applied to vacuum breakdown in the high constant voltage range.

### PULSED MICROWAVE VOLTAGES

Some studies have recently been made by the linear accelerator group at M.I.T. of the insulation of microwave voltages in high vacuum.[12] In a copper cylindrical cavity, 3000 mc voltages in pulses 2 $\mu$sec. in duration and calculated at $2\times10^6$ volts peak were insulated across the 5-cm gap. The pulse nature and high frequency of these insulation tests is in favor of both high voltages and gradients since the time and energy available to accomplish breakdown is limited. It would seem that positive ions can have relatively less influence with this type of voltage and that the breakdown mechanism would depend on high field and photoelectric emission of electrons from both electrode surfaces viewed as alternate cathodes.

### POSSIBILITIES FOR REDUCING SURFACE BREAKDOWN COEFFICIENTS

The striking dependence of secondary electron emission on the electrode material illustrates the way in which the breakdown coefficients $A$ and $B$ may also depend on materials, and hence the need of quantitative information on these factors. When these breakdown coefficients are sufficiently known, it should be possible to begin to design and predict the insulation strength of electrodes in vacuum. Moreover, it may prove possible to improve substantially the insulating strength between electrodes in vacuum by various measures which reduce the electrode coefficients. Various possible procedures are suggested, including the use of composite electrode surfaces in which the surface layer is of low atomic number and, therefore, inefficient as a source of heavy secondary particles and x-rays, the use of highly porous surfaces and of grids or insulating films to repress secondaries, the local distortion of the surface field so as to delay the focusing effects of a particle interchange, the magnetic deflection of the interelectrode electrons, and the more complete conditioning of the electrode surfaces. The experimental determination of surface coefficients under these various conditions can in principle be performed in a more quantitative way than is generally possible in experiments on material insulating media. Such studies should not only widen our understanding of the essential factors in vacuum breakdown, but they may extend insulation strength of high vacuum into the region of usefulness for electrical power engineering and various scientific purposes.[13]

### ACKNOWLEDGMENT

It is a pleasure to acknowledge the effective assistance of J. R. Maull in making some of the measurements reported herein.

---

[12] Halpern, Eberhardt, Rapuano, and Slater, Phys. Rev. 69, 688 (1946).

[13] J. G. Trump, Doctorate Thesis, M.I.T., E. E. Dept. (1933).

---

### Calendar of Meetings

**March**
22–27 American Society for Metals, Oakland, California
28 Institute of Aeronautical Sciences, Cleveland, Ohio

**April**
4–5 American Physical Society, Southeastern Section, Salisbury, North Carolina
9–11 Society of Automotive Engineers, New York, New York (Aeronautic Meeting)
14–18 American Chemical Society, Atlantic City, New Jersey
16–18 Society of Automotive Engineers, Chicago, Illinois (Transportation Meeting)
19 Institute of Radio Engineers, Chicago, Illinois
21–25 American Ceramic Society, Inc., Atlantic City, New Jersey
21–25 Society of Motion Picture Engineers, Chicago, Illinois
22–23 American Geophysical Union, Portland, Oregon (Section of Hydrology)
23–26 American Society of Civil Engineers, Phoenix, Arizona
24–26 American Philosophical Society, Philadelphia, Pennsylvania
25–26 American Mathematical Society, Chicago, Illinois
25–26 American Mathematical Society, New York, New York
26 American Mathematical Society, Stanford University, California
28–30 American Geophysical Union, Washington, D. C.
28–30 National Academy of Sciences, Washington, D. C.

**May**
1–3 American Physical Society, Washington, D. C.
8–10 Acoustical Society of America, New York, New York
17 New England Radio Engineering Meeting, Cambridge, Massachusetts
26–27 Institute of Aeronautical Sciences, Detroit, Michigan

Journal
of
Applied Physics

Volume 18, Number 4 April, 1947

# Preparation and Uses of Silica Replicas in Electron Microscopy

CHARLES H. GEROULD
*The Dow Chemical Company, Midland, Michigan*

(Received November 21, 1946)

A detailed description of the preparation of silica replicas and substrates and their many and varied uses is presented with photographs and electron micrographs. A method of preparing silica replicas of specimens which cannot be subjected to the temperatures and pressures of the ordinary technique is described. It consists of applying a polystyrene lacquer to the specimen surface in place of the conventional molding. The question of "up and down" in stereo pictures of silica replicas is definitely settled by adding a metal oxide smoke (MgO) to the polystyrene replica before the silica evaporation. A successful powder dispersing technique consists of dispersing the powder in an ethyl cellulose lacquer and depositing upon a silica substrate. Several other techniques are also included.

## INTRODUCTION

WITHIN the last two years the field of electron microscopy has expanded many fold resulting in a great need for the development of new and improved electron microscopic techniques as well as the optimum use of those available. The great diversity of the recent entrants into the microscopic field has made an even greater need for literature on techniques which are of general application.

The polystyrene-silica replica technique of Heidenreich and Peck[1] is one which has succeeded in solidly establishing itself and is now being widely used on many and varied research problems involving the electron microscope. Among these we might mention the application of silica replicas to such metallurgical specimens as steels, copper alloys, and magnesium alloys.[2–5] We might mention also their application to nonmetallic materials such as teeth,[6] bones, minerals, glasses, ceramics, resins, woods, and textile fibers.[7] In addition, we must include the use of structureless silica substrates[8] as supports for many different types of electron microscope specimens such as bacteria, viruses, powders, emulsions, metallic oxide smokes, thin evaporated films, microtome sections, etc.

These many and varied uses of silica replicas have resulted in a demand for more explicit information on their preparation and uses. It is the purpose of this paper to fill this need and at the same time make available to the field some of the more recently developed, and as yet unpublished, microscope techniques in which silica replicas are an essential part.

These new techniques include a new method

---

[1] R. D. Heidenreich and V. G. Peck, J. App. Phys. 14, 23 (1943).

[2] C. S. Barrett, J. App. Phys. 15, 691 (1944).

[3] R. D. Heidenreich, SAE J. (trans) 53, 588 (1945).

[4] R. D. Heidenreich, L. Sturkey, and H. L. Woods, J. App. Phys. 17, 127 (1946).

[5] R. D. Heidenreich, C. H. Gerould, and R. E. McNulty, Tech. Pub. 1979, Metals Tech. 13 (April, 1946).

[6] C. H. Gerould, J. Dent. Res. 23, 239 (1944).

[7] R. B. Barnes, C. J. Burton, and R. G. Scott, J. App. Phys. 16, 730 (1945).

[8] R. D. Heidenreich, J. Opt. Soc. Am. 35, 139 (1945).

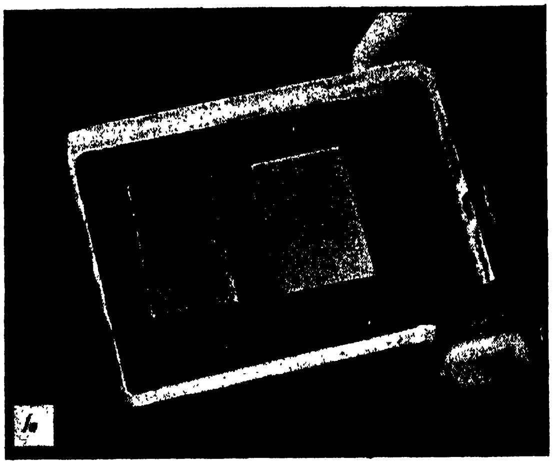


FIG. 1a. Block of clear polystyrene containing two specimens of a magnesium alloy sheet whose surfaces have been covered with a corrosion resistant coating. Silica replicas are to be prepared of this coating. The molding was formed from molding granules at a temperature of 160°C and a pressure of about 1500 lb./sq. in.

FIG. 1b. Polystyrene impression on left and coated metal specimen on right. Mechanical separation has been accomplished by sawing away the excess polystyrene from the edges of the specimen and then pulling the two surfaces apart. Specimens may be dissolved with acids or alkalies if necessary.

of preparing replicas for stereoscopy which makes it possible to settle definitely the question of "up and down" in stereo pictures, a procedure for preparing silica substrates, an ethyl cellulose lacquer particle dispersing technique for use with such substrates, a method of preparing silica replicas utilizing a room temperature produced polystyrene impression and several other miscellaneous techniques.

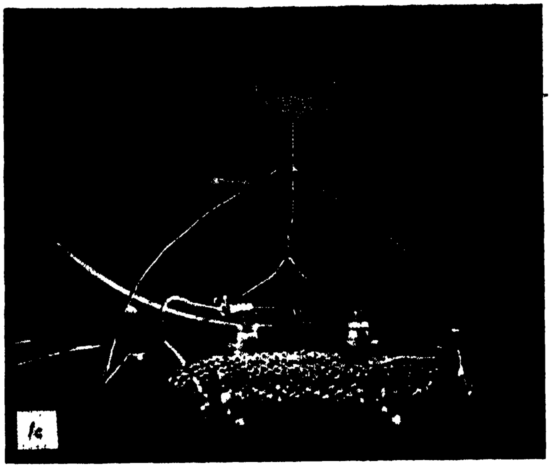


FIG. 1c. Polystyrene impression of specimen mounted 7 to 10 cm above conical tungsten filament containing approximately 1 mg of silica in the form of quartz.

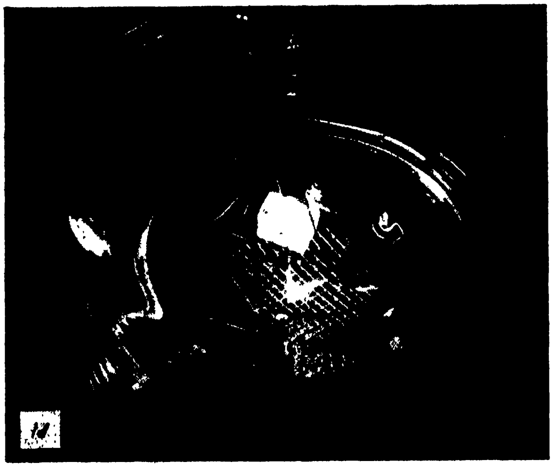


FIG. 1d. Actual evaporation of silica with filament incandescent under a vacuum of about $10^{-4}$ mm of Hg. Silica is completely evaporated in 5 to 15 seconds, part of which condenses on surface of polystyrene to give a continuous film of approximately 200A thickness.

All micrographs were obtained with either the RCA type B or U electron microscope.

## PREPARATION OF SILICA REPLICAS

The cleanliness of the specimen surface is one of the most important points to be considered in preparing silica replicas. In the case of an etched metal the prepared surface must be free of reaction products, grease, stains, oxide, or hydroxide films or any material which may either adhere to the replica or prevent it from reproducing the true metal surface. When working with non-metallic specimens, the surface must likewise be free of any foreign material which would contaminate the replica. It is our belief that many of the difficulties met in preparing silica replicas by the polystyrene-silica technique arises from improperly prepared surfaces, mainly dirty surfaces.

The polystyrene-silica technique can best be illustrated by a series of photographs of the various steps involved in preparing the polystyrene impression, evaporating the silica upon the plastic, and removing the silica replica film. These various steps are described in detail in order that one may experimentally follow the technique and obtain good replicas with a minimum of difficulty.

The first step, after the specimen has been given a suitable surface preparation, is to mold polystyrene against the surface. This is accomplished quite simply by imbedding the specimen in polystyrene as illustrated in Fig. 1a. Polystyrene molding granules of medium viscosity or molecular weight* are best for this.

The molding may be accomplished in an ordinary electrically heated metallographic specimen mounting press or a steam-heated compression molding die. The specimen is placed in the molding die with the prepared surface upward and sufficient molding granules added to give a final clear molding at least ¼ inch thicker than the specimen. The actual molding is accomplished by heating the molding granules to a temperature of 160°C, care being taken that no pressure is applied until they are thoroughly softened. Otherwise they may be pressed against the specimen surface causing deformation. After the granules have had a chance to soften thoroughly (5 to 10 minutes usually is required), pressure is applied up to about 1500 lb./sq. in. and the cooling of the die commenced at once. The pressure is maintained until the molding has cooled below 80°C since release of the pressure at higher temperatures may result in the formation of bubbles in the molding.

The specimen is now separated from the plastic impression by first sawing away the excess material from the edges of the specimen and then separating the two, mechanically if possible. Figure 1b shows a magnesium alloy specimen which has just been mechanically separated or stripped from the polystyrene.

If mechanical separation is impossible without fracturing or otherwise distorting the polystyrene, chemical reagents may be used to dissolve the specimen. Mineral acids or strong caustic solutions may be used on most specimens without damage to the plastic impression.

Figures 1c and 1d show the next step in which silica is evaporated in vacuum upon the surface of the plastic impression. The molding is mounted 7 to 10 cm above a spiral tungsten filament (cone shaped of about 5-mm length with a 35° angle) containing about 1 mg of clear quartz in the form of small slivers weighing roughly 0.2 mg each. The filament of the type shown in the photographs is heated to incandescence for a period of 5 to 15 seconds, using 25 to 30 amp. at 10 to 15 volts. The silica is completely evaporated in a few seconds, providing the individual splinters of silica are not too large and the vacuum is $10^{-4}$ mm of Hg or better.

The tungsten filament used in this evaporation is prepared by winding 20-mil tungsten at a dull red heat on a tapered mandrel having about a 35° included angle. The quartz used may be prepared quite easily by breaking up a natural clear quartz crystal by means of mortar and pestle and then screening to obtain particles of the proper size. Other laboratories have successfully used cloudy quartz, fused silica, or Santocel.

The block of polystyrene supporting the silica film is now prepared for the removal of the film. The sides and bottom are first scraped thoroughly to remove all undesirable silica film and any dirt which may contaminate the solvents. The silica film is scored into slightly over ⅛" squares and the polystyrene block lowered to the bottom of a dish of high purity ethyl bromide and benzene with the replica side up as shown in Fig. 2a. The block will stick to the bottom of the dish as the plastic softens. About 10 percent of benzene is added to the ethyl bromide to decrease its volatility and thus prevent moisture condensation upon the silica films as they are removed from the solvent. Such moisture condensation is invariably accompanied by precipitation of the polystyrene. Ethyl bromide is used for this step

---

* Such a material is available commercially as Styron No. 379-clear.

FIG. 2a. Silica film has been scored into squares and the whole polystyrene block is about to be immersed in ethyl bromide plus 10 percent benzene.

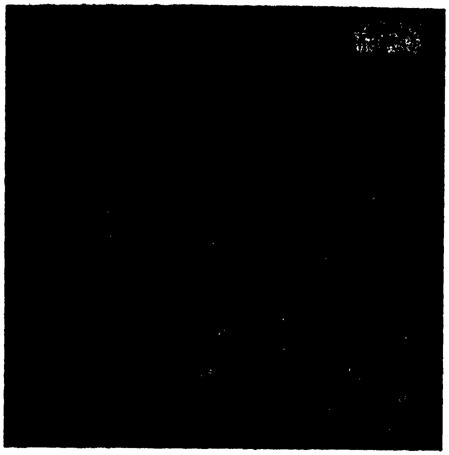
FIG. 2c. Silica films are being transferred on a piece of 200-mesh stainless steel screen to a rinse consisting of ethyl bromide plus a few percent of iso-amyl alcohol.

FIG. 2b. Silica films are seen as they start floating off from the polystyrene block stuck to the bottom of the dish of ethyl bromide and benzene. Silica films generally come free from the polystyrene 1 to 5 min. after immersion in the solvents.

FIG. 2d. Catching the silica films on a standard electron microscope specimen screen after they have washed in the solvent rinse for at least 5 min. After catching, the screens are blotted on a filter paper and viewed directly in the electron microscope.

in the replica technique for the following reasons: 1. It is an excellent solvent for polystyrene; 2. its density is sufficiently high to cause the films to float about in the solvent; 3. its great volatility tends to produce convection currents which aid in washing the films free from polystyrene; 4. its refractive index is such that the silica films are generally readily visible in the solvent.

The silica films will generally free themselves from the polystyrene 1 to 5 minutes after immersion. Side illumination with a dark red background is of great help in seeing them. Figure 2b shows the films as they come free from the polystyrene. They are transferred almost at once to a rinse containing ethyl bromide plus a few percent of iso-amyl alcohol added to reduce the solvent's volatility. Benzene is not used because an effective concentration is probably upwards of 10 percent. Transfer of the silica films to the rinse is easily accomplished by using a piece of 200-mesh stainless steel screen as shown in Fig.



Courtesy A.I.M.E.

FIG. 3. Silica replica of magnesium alloy showing twinning and fine globular precipitation, 4000×.

FIG. 6. Silica replica of cleaved enamel from human tooth, 3500×.

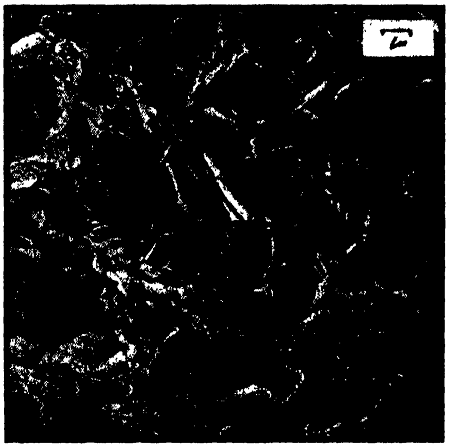


FIG 4. Silica replica of ground surface of quartz crystal, 4000×.

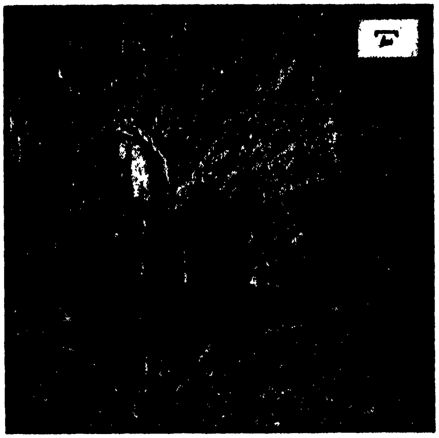


FIG. 7. Silica replica of polished and acid-etched internal section of rabbit bone, 3000×.

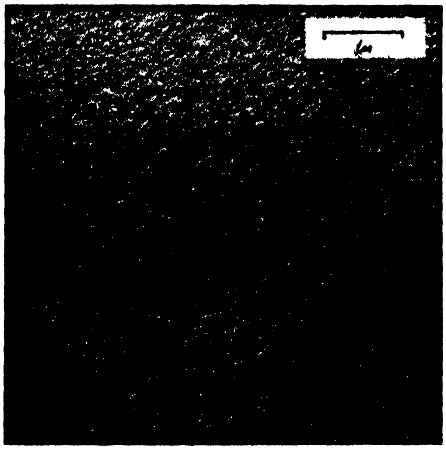


FIG. 5. Silica replica of glass etched with HF, 10,500×.

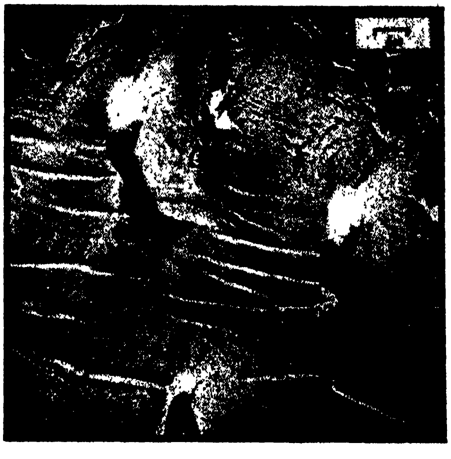
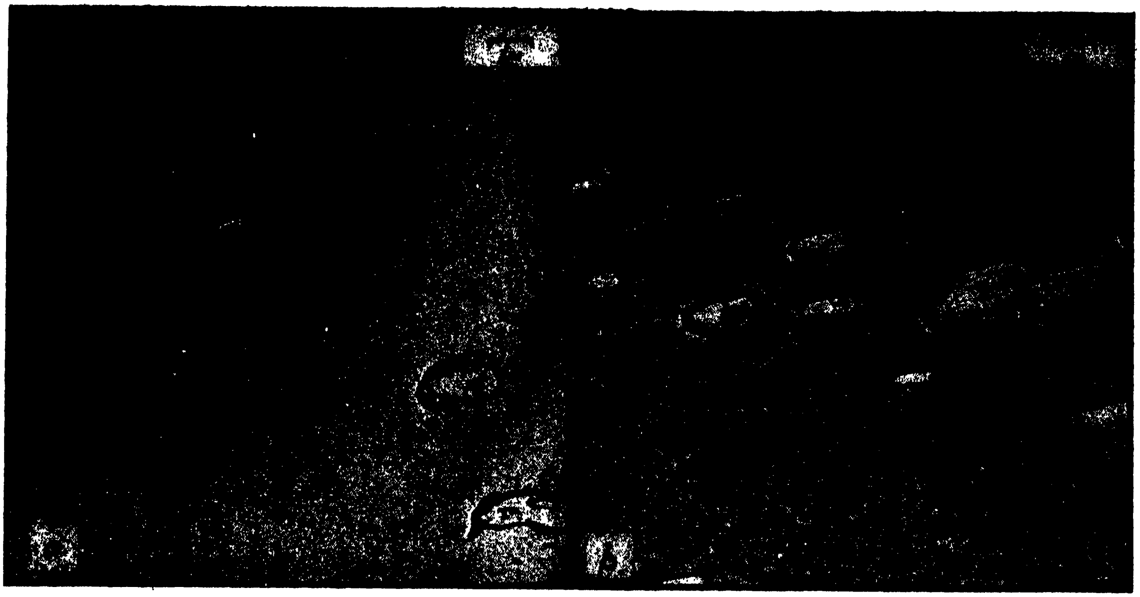


FIG. 8. Silica replica of maple wood cleavage, 4500×.



Courtesy A.I.M.E.

FIG. 9. Comparison between silica replicas of a magnesium alloy prepared by means of: (a) a polystyrene molding using both heat and pressure, 6000×. (b) a polystyrene lacquer film using no heat or pressure, 6000×.

2c. The films readily loose themselves from the screen when it is lowered into the rinse provided they are not allowed to dry during transfer.

The silica films are allowed to wash at least five minutes and then caught on the standard microscope specimen screens as shown in Fig. 2d. The screen with the films are dried by blotting on a filter paper. Such films are believed to have a thickness of about 200A.

Figures 3–8 show examples of silica replicas as applied to a magnesium alloy, quartz crystal, etched glass, tooth, bone, and wood.

One often finds in replica work a need for obtaining silica films from specific areas of a specimen. This is not difficult since the polystyrene molding with the silica film may be examined quite readily with the light microscope at magnifications up to 1000× using illumination through the specimen. Regions which are of particular interest may be outlined readily by scribing the surface with a needle point and then scraping off all the silica film except that of particular interest. This film may then be removed in the usual manner. Such a technique is often found very useful in studying interfaces between dissimilar materials.

As has been mentioned before, many of the difficulties experienced when preparing silica replicas arise from improperly prepared surfaces. Dirty surfaces often give silica replica films which tend to roll or fold up and which may refuse to straighten out by the customary method of "teasing" the films with a specimen screen as they float in the rinse. Another difficulty often experienced in preparing silica replicas is fragility of the films caused by lack of appreciable fine structure in the original specimen. It appears that fine detail in a specimen adds considerably to the mechanical strength of silica replicas.

On the other hand, silica replicas of single phase or electrolytically polished metals, glass, or other microscopically smooth surfaces are very often extremely fragile. Care must be taken to handle these films delicately and even then part of the film may break up on the specimen screen.

Occasionally one wishes to prepare replicas of surfaces which for some reason cannot be subjected to the temperatures and pressures necessary to obtain a polystyrene impression. In such cases an alternate method is used in which a polystyrene lacquer is applied to the specimen to give a heavy coating, stripped off and then used to prepare silica films in the ordinary manner. This technique has given quite reliable results in practice using a 1 percent solution of polystyrene in benzene for the initial application and then backing this film up with several applications of a more viscous lacquer of 2 to 3 percent polystyrene in benzene. By means of the more concentrated solution one can build the original film up to the point where it may be mechanically stripped from the specimen or the specimen dissolved if it is felt to be advisable. Figures 9a and b show a comparison between the conventional method and this polystyrene lacquer technique.

The resolution possible with this technique appears to be very good though very likely not equal to that of the pressure molding technique whose resolution is apparently better than 100A. The resolution of this "cold" replica technique can probably be improved, as suggested by Matheson,* by vacuum boiling of the specimen in the lacquer as a means of outgassing the surface. The initial lacquer coating can then be built up as previously described.

## STEREOSCOPY

The use of stereoscopy as an electron microscope technique has become of quite general application and is held to be invaluable in interpreting electron micrographs by many microscopists. The stereoscopic technique is especially well suited for use with silica replicas.[9] However, one of the difficulties experienced in such application has been the impossibility of definitely establishing the "up and down" of the stereo pictures unless something was known about the topography of the original specimen surface.

A very simple procedure is now being used to remove this uncertainty. This consists of adding a metal oxide smoke to a known side of the silica replica film and using this oxide smoke to determine the correct viewing arrangement. This is best accomplished by depositing the oxide smoke on the polystyrene impression of the specimen surface. Magnesium oxide smoke works very successfully, an even distribution being obtained by wafting the molding through the smoke of burning magnesium. Zinc oxide smoke can also be used but is much more difficult to distribute evenly.

Silica is evaporated upon the polystyrene impression and the films removed in the usual manner carrying the oxide smoke with them,

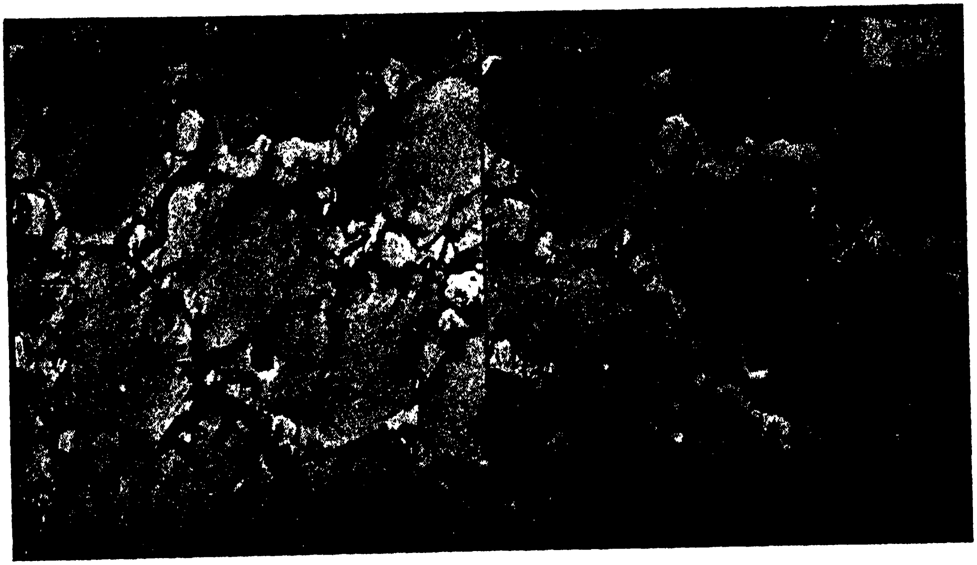


FIG. 10. Stereo pair from a fine grain magnesium alloy showing the addition of magnesium oxide smoke (indicated by arrows) to the replica film. The question of "up and down" is answered by viewing the stereo pair arranged so the oxide cubes appear below the surface. Compound precipitation can be observed projecting upward in relief from the grain boundaries. Magnification 6000X.

---

* L. A. Matheson and R. D. Heidenreich, "Magnification calibration of electron microscopes," Presented at Chicago meeting of E.M.S.A. (November, 1944).

[9] R. D. Heidenreich and L. A. Matheson, J. App. Phys. 15, 423 (1944).

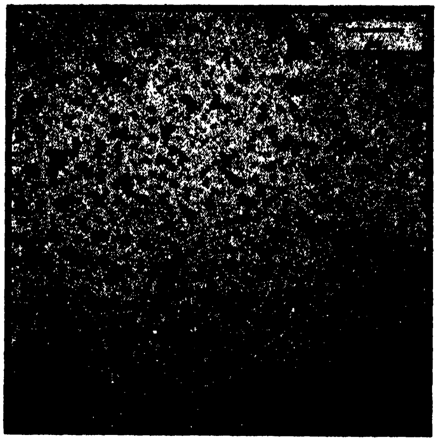


FIG. 11. Silica smoke dispersed in ethyl cellulose on a silica substrate, 7500×.

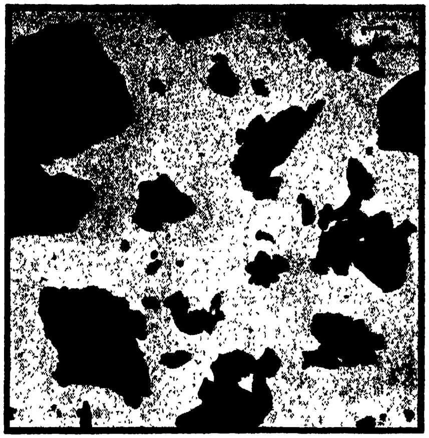


FIG. 12. $CaCO_3$ dispersed in ethyl cellulose on a silica substrate, 4500×.

attached to the side reproducing the specimen's surface structure. Electron micrograph stereo pairs of these films are then obtained and viewed through the stereoscope with the pictures arranged so the oxide smoke appears below the surface, the proper arrangement. By this method the question of "up and down" is definitely settled. Figure 10 shows a stereo pair of a magnesium surface with MgO smoke on the replica.

## SILICA SUBSTRATES AND A POWDER DISPERSING TECHNIQUE

Structureless silica substrates have been found very useful as supports for a variety of electron microscope specimens. They can be prepared having a thickness of from <100 to 200A and are completely structureless and without holes or fibers. The films are structurally strong as compared to organic films, and may be subjected to very intense electron beams with a minimum of electron scattering, "charging up," and breakage. Microscope specimens may be heat treated at elevated temperatures for short periods of time while supported on silica substrates. They also can be successfully used as the support for specimens for electron diffraction transmission study.

These substrates are prepared by dipping a clean and dry microscope slide in a 2–5 percent solution of polystyrene in benzene and allowing to dry. The silica film is formed upon the lacquer film by the same technique as used for replicas except that 10 or 20 percent less silica is generally used. The films are removed in the same manner as other silica replica films and caught on standard electron microscope specimen screens. However, the silica films separate from the thin polystyrene film quite slowly, ½ hour to several hours usually being required.

Difficulties are occasionally met in preparing these films and they seem to be largely caused by improper thickness of the polystyrene film. For some inexplicable reason, if this film is too thick, the silica films are fragile, and if too thin, the silica films come loose from the lacquered slide with great difficulty or may not come loose at all.

FIG. 13. Polystyrene emulsion dispersed on a silica substrate, 7500×.

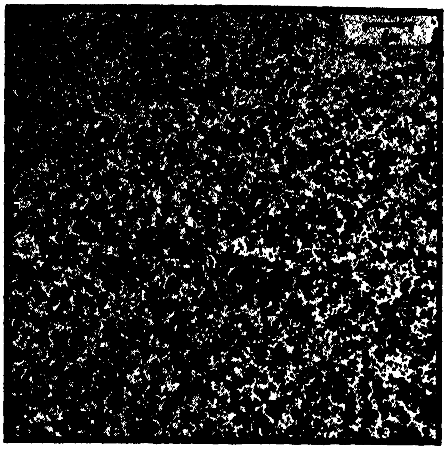


FIG. 14. Magnesium-aluminum alloy evaporated in vacuum and condensed on a silica substrate, 6500×.

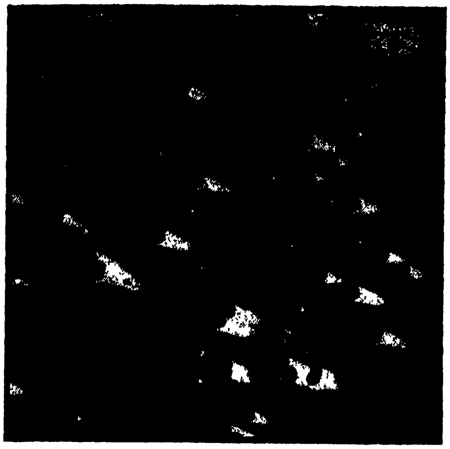


FIG. 16. Lampblack dispersed in ethyl cellulose on a silica substrate and shadowed with chromium, 3500×.

It appears that a 2–5 percent lacquer solution gives a film of optimum quality.

Precautions must also be taken to see that the lacquer is relatively moisture-free and the glass slide dry. Failure to observe these precautions may result in a cloudy lacquer film and substrates which show bubbles.

Many different particle dispersing techniques have been published[10, 11] which make it possible to determine accurately particle size and shape. Most of these techniques can be successfully used with silica substrates with improved results. A powder dispersing technique has been in use here for several years which has given excellent results on a wide variety of powdered materials ranging in size from <0.01 micron to several microns. This technique consists of dispersing the powder in an ethyl cellulose lacquer as a thick slurry, diluting and adding to the silica substrate. This is accomplished in practice by placing a few milligrams of the powder in a mortar and adding a drop or two or an ethyl cellulose lacquer consisting of 2 percent ethyl cellulose in 70 percent amyl acetate and 30 percent ethyl alcohol. The mixture is lightly ground with mortar and pestle to a thick slurry or paste during which time the dispersion of the particles actually takes place. Amyl acetate is then added slowly at first and the concentrated dispersion diluted to the proper concentration for adding to the silica substrate on a specimen screen. Tests have indicated that the effect of the grinding action upon the particle size and shape is negligible in specimens having a particle size of less than 5 microns.

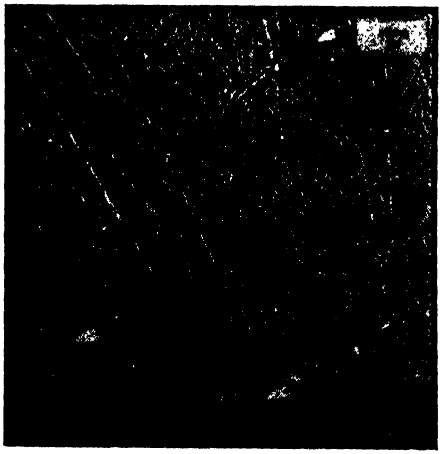


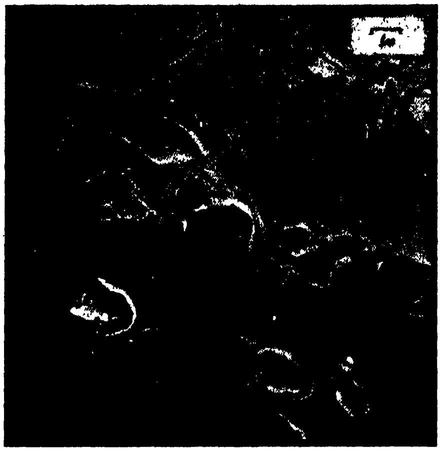
FIG. 15. Silica replica of a polished steel surface shadowed with chromium, 3500×.

Ability to judge the proper quantities and dilutions is readily obtained with a little experience. Generally the slurry dilution is about 1:50. Quantities should be varied so that the minimum thickness of dried ethyl cellulose is used, i.e., just sufficient to keep the particles firmly attached to the silica film. Figures 11 and 12 show typical particle dispersions obtained by this technique.

---

[10] M. L. Fuller, P. G. Brubaker, and R. W. Berger, J. App. Phys. 15, 201 (1944).

[11] A. M. Cravath, A. E. Smith, J. R. Vinograd, and J. N. Wilson, J. App. Phys. 17, 309 (1946).



Courtesy A.I.M.E.

FIG. 17. Silica replica of a magnesium-nickel alloy showing not only a replica of the globular $Mg_2Ni$ particles but also some of the actual particles themselves on the silica replica film. The faithfulness of the technique in reproducing the size and shape of the particles is evident. Figure 3 also shows this same phenomenon on a smaller scale. Magnification 4000×

Silica substrates may also be used for supporting a variety of other types of specimens as illustrated by Figs. 13 and 14 showing an emulsion and an evaporated magnesium alloy. Emulsions are readily dispersed by diluting to the proper concentration and then adding a drop of the liquid to the substrate on a specimen screen.

## MISCELLANEOUS USES AND TECHNIQUES

Silica replicas and substrates have also been found to be useful in the field of heavy metal "shadow-casting."[12] This technique is probably not of great value on silica replicas for the purpose of improving contrast since silica replicas normally show considerable contrast. However, the technique has definite possibilities as a means of making elevation determinations without the use of stereoscopy.

Shadowing of silica replica films is probably best accomplished by evaporating the heavy metal upon the silica film while it is on the surface of the polystyrene molding. The films are then removed in the usual manner. Figure 15 shows an example of the "shadow-casting" of a polished steel specimen.

Shadowing of various electron microscope specimens supported on silica substrates is of definite value both from the point of view of improving contrast and of making elevation determinations. This technique is probably best accomplished by applying the polystyrene lacquer film to a glass slide, evaporating the silica, adding the specimen, "shadow-casting," and then removing the silica film in the usual manner. It seems advisable to use such a procedure rather than attempt the shadowing of the substrate and specimen while supported by the specimen screen. In this case the silica film may be uneven, making it a very poor target for accurate "shadow-casting." Figure 16 shows a specimen of lampblack "shadow-cast" in this above recommended manner.

A very interesting and often valuable use of silica replicas lies in the isolating of impurities and certain second phase elements or compounds from some metals. This occurs occasionally in metals which contain one or more elements or compounds which are present as a second phase and which are relatively insoluble in the etchant used during the surface preparation. Such constituents of the metal are often either left lying on the metal surface after etching or are protruding from the surface and only loosely attached to the metal. When polystyrene is molded against the metal surface and separated, these particular constituents are imbedded in the polystyrene and removed with it. When the silica evaporation is completed and the polystyrene dissolved, as the silica films are removed, the particles are often transferred to the silica replica and attached to it. As a result we find that in the electron microscope we see not only a replica of certain second phase constituents in the metal but also some of these particular constituents actually attached to the replica film. Figures 3 and 17 show examples of this phenomenon. Such materials can often be identified by transmission electron diffraction.

Occasionally one finds an electron microscope specimen in particle form which tends to sublime when subjected to the high vacuum and heat of electron bombardment in the microscope. This sublimation may be appreciably decreased by supporting the specimen on a silica substrate

---

[12] R. C. Williams and R. W. G. Wyckoff, J. App. Phys. 15, 712 (1944).

and then evaporating more silica over the surface of the specimen. Sulfur is a material which may be treated in this manner.

### CONCLUSIONS

1. Silica replicas of a variety of bulk materials such as metals, minerals, glasses, ceramics, resins, woods, etc., can be successfully prepared using a heat and pressure molded polystyrene impression plus evaporated silica.

2. Silica replicas can be successfully prepared from bulk materials which cannot tolerate either the temperatures or pressures of (1), by applying a heavy polystyrene lacquer, removing and evaporating silica upon the lacquer impression.

3. The question of "up and down" in stereo pictures of silica replicas can definitely be established by depositing a metal oxide smoke on the polystyrene impression before evaporation of the silica.

4. Structureless silica substrates are well suited as the supporting membranes for specimens such as powders, emulsions, metallic oxide smokes, thin evaporated films, bacteria, viruses, microtome sections, etc.

5. Excellent dispersions of powders from <0.01 to several microns in diameter may be obtained on silica substrates using an ethyl cellulose lacquer dispersing technique.

6. Heavy metal "shadow-casting" can be effectively applied to both silica replicas and silica substrates.

7. Isolation of metallic phases directly on silica replicas of metal surfaces is often useful in minutely examining and identifying these constituents.

### ACKNOWLEDGMENT

The author wishes to express his appreciation to R. D. Heidenreich of the Bell Telephone Laboratories and H. L. Woods of the Henry Ford Hospital who developed several of the techniques described in this paper while employed at The Dow Chemical Company.

---

# The X-Ray Storage Properties of the Infra-Red Storage Phosphor and Application to Radiography

O. E. Berg and H. F. Kaiser
*Naval Research Laboratory, Washington, D. C.*

(Received November 4, 1946)

An application of the characteristics of infra-red storage phosphors to radiography is described, whereby the expense and trouble of x-ray film processing may be partially or completely eliminated. The properties of these materials to store x-ray energy for periods of time and release that energy upon exposure to infra-red radiation have proved practical for either temporary and permanent recording of the latent x-ray image. One distinct feature of this method is the complete removal of harmful x-radiation to the technician. The possibility of a new field of radiography in "flash fluoroscopy" is also suggested by the authors. Graphs and charts exhibiting the characteristic behavior of infra-red phosphors to x-ray energy are included as well as actual photographs and radiographs comparing radiography with phosphorography.

The image storage phosphor was a research development required in war work as a practical and sensitive means for the detection of infra-red radiation. In the course of research[1] carried on to produce such phosphors it was found possible to produce phosphors which exhibited storage properties not only for visible light, but also for cathode rays, and x-rays. The optical characteristics of such materials have been described in papers of Roland Ward[2] and of N. F. Miller and C. E. Barnett.[3]

---

[1] Contract OEMSR 982, OSRD Report 16.5—119 (N5643).

[2] R. Ward, "Preparation and properties of infra-red sensitive strontium selenide and sulfide-selenide phosphor," J. Opt. Soc. Am. 36, 351(A) (1946).

[3] N. F. Miller and C. E. Barnett, "Infra-red sensitive phosphors from zinc and cadmium sulfides," J. Opt. Soc. Am. 36, 352(A) (1946).

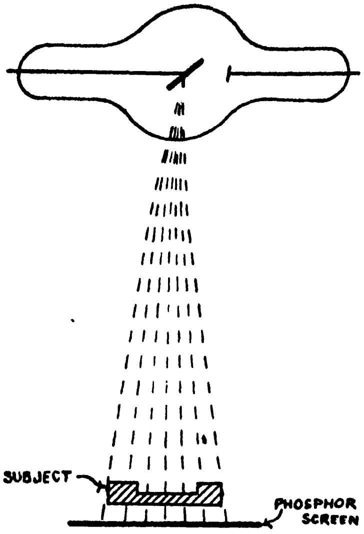


FIG. 1A.

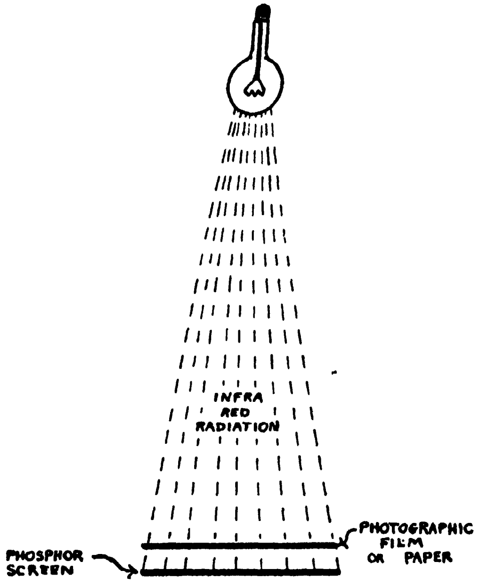


FIG. 1B.

Since no attempt had been made to study these materials for x-ray energy storage characteristics, the authors have carried out a series of experiments on materials kindly furnished by the Brooklyn Polytechnic for this purpose. The ability of these materials to save x-ray energy suggests their application to radiography as a possible substitute for the photographic emulsion for the retention of a latent image.

The ability of these materials to store radiant energy is not surprising in view of the fact that all phosphors exhibit such properties when cooled to liquid air temperature, exposed to an energy source to fill the electron traps and then warmed up, whereupon the freeing of electrons from the traps is indicated by bursts of light.[4]

Phosphors may be produced which store energy from x-rays, beta-rays, gamma-rays, and visible light and which hold this energy for indefinite periods of time. The energy is released, in the form of light varying from red to ultraviolet, by infra-red radiation with speed of release governed by intensity of the infra-red radiation. Figure 1A and B shows how these effects may be employed in radiography. The x-rays pass through the subject into the phosphor plate or screen, and by differential intensity record in detail thereon producing a latent image. This latent image is quite stable for many phosphors.

Then, in an adjoining room, away from harmful x-radiation, and at the convenience of the operator, the image is made visible to the eye by irradiating the plate with infra-red radiation. If a particular area of the image is of special interest to the technician, he may expose it separately with an infra-red spotlight (photographic masking technique is also applicable here). The length of viewing time of a phosphor plate may be determined from its energy decay rate graph shown in Fig. 2A which indicates the intensity of the released light as a function of time and intensity of the infra-red source.

When plotting light units against log time as in Fig. 2B, we get several linear portions indicating that the decay of luminescence is not of simple exponential character, but occurs in a set of several such processes. Light values are given as arbitrary light units measured by a 931A photo-cell multiplier with voltage regulators.

If a permanent record of the x-ray image is desired, it is an inexpensive and a simple task to expose the phosphor image directly to a sensitized paper (sensitive in the color region of the particular phosphor in use). Such a process eliminates the need for double emulsion x-ray film.

A 4"×5" experimental phosphor viewing plate was prepared according to the following procedure:

---

[4] "Characteristics of luminescent materials for cathode-ray tubes," J. Inst. Elec. Eng. 92, Part III, No. 20, 300 (December, 1945).

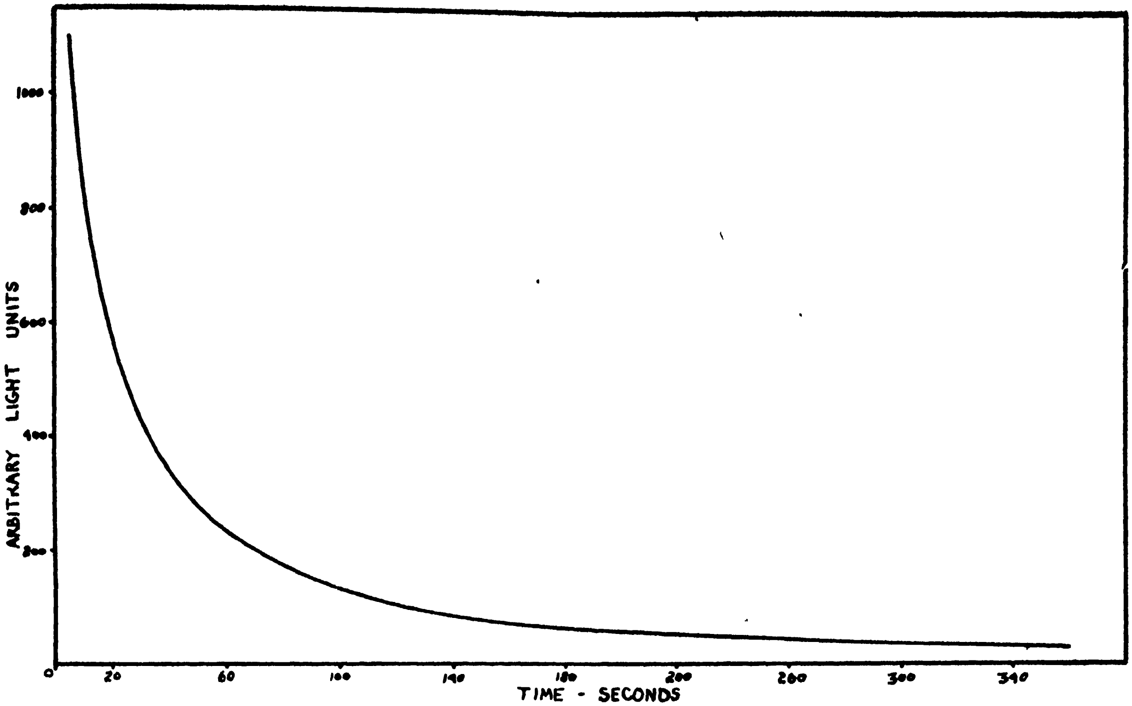


FIG. 2A.

While it might have been desirable to apply the phosphor to a thin cardboard base as for a fluorescent screen, because of the high processing temperatures involved, it was found necessary and advantageous to use a $\frac{1}{32}''$ steel plate with a thick silver plating. This base plate was not affected by high temperatures and its silver coating prevented chipping or peeling of the phosphor during experimentation.

Furthermore, silver plated steel is preferable to a paper base in that it serves as an absorbing medium for transmitted x-radiation after the x-rays have transversed the subject and the phosphor, thereby reducing scatter.

An emulsified mixture of phosphor, ethylene dichloride, and jasonite (plastic binder) was applied with a small artist's spray gun, and dried at 850°C to remove excess binder and render the phosphor active.[5] After a few experiments, it was found that a .020-inch coating of the phosphor was sufficiently thick to produce peak luminescence.

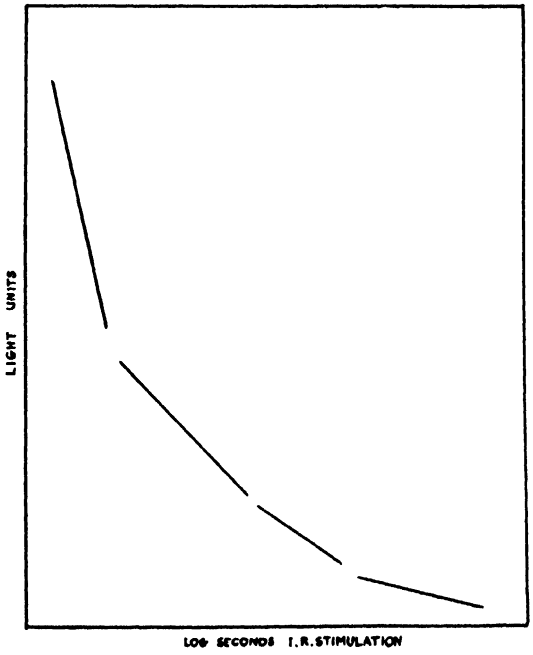


FIG. 2B.

Test exposures were made on the resultant plate at various kilovoltages and exposures. Figures 3 and 4 are comparative results of an x-ray film image and a phosphor image, respectively, of a 3-inch ammeter. It is an indication

---

[5] For further information on phosphor coating application see OSRD Report No. 5643.

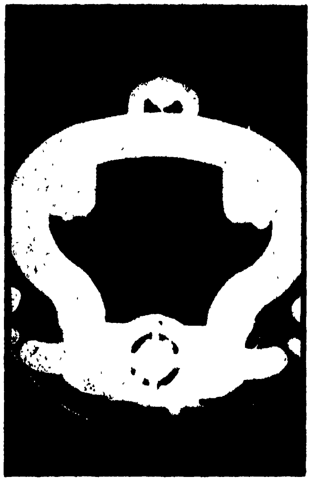


FIG. 3. Radiogram of 3 inch meter. X-ray exposure made on fine grain industrial x-ray film.

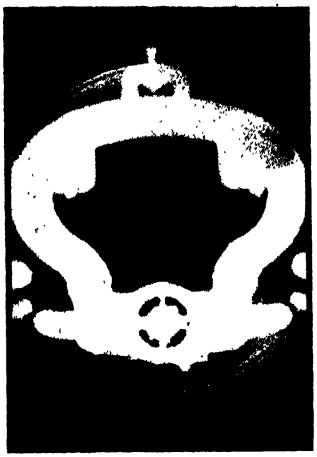


FIG. 4. Phosphorogram of 3-inch meter. Dark areas indicate uneven phosphor coating.

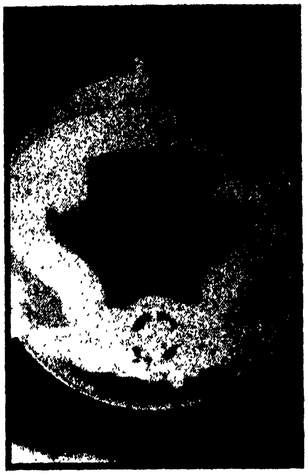


FIG. 5. Secondary film exposure to phosphor image of meter. Notice increased latitude and preservation of detail.

of the definition and latitude obtainable in a "phosphorogram."

An interesting and advantageous feature concerning latitude in this method of exposure is revealed in Fig. 5, another photograph of the same exposure after some diminution by previous

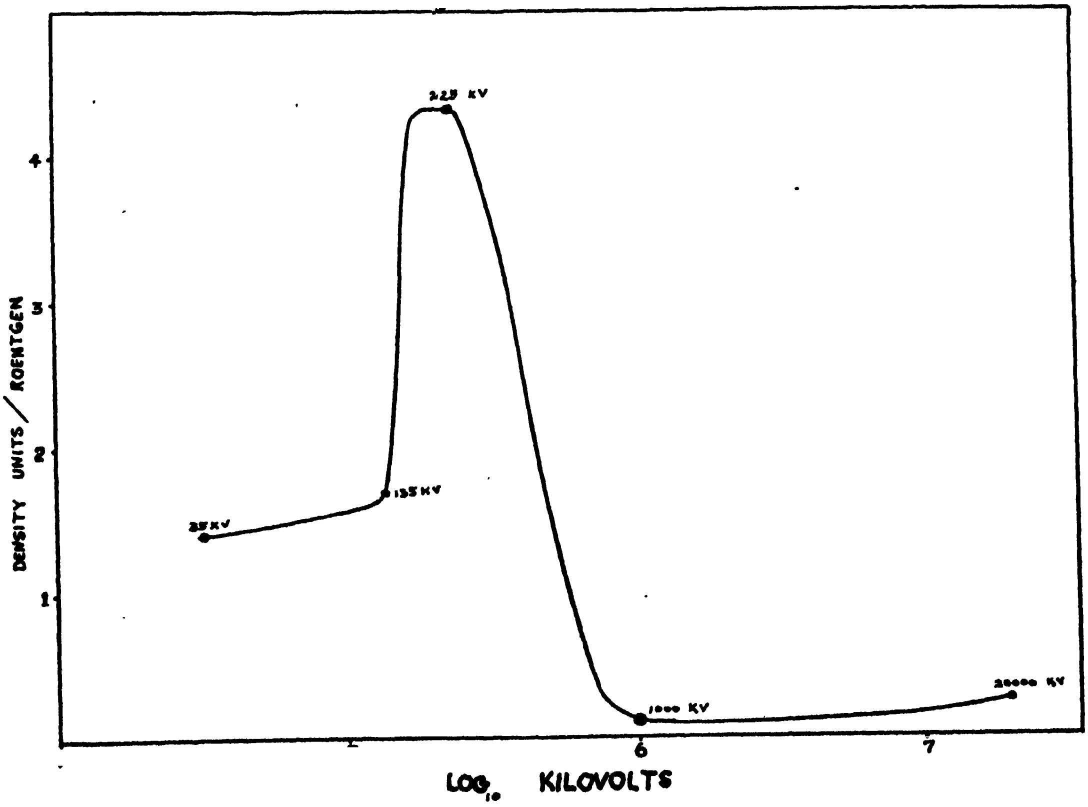


FIG. 6. The storage sensitivity curve in the x-ray region of the spectrum for one particular infra-red phosphor.

infra-red exposure. The definition is as sharp as before and areas that were overexposed (too brilliant) in the previous photographic exposure now appear in readable detail. The latitude of a phosphorogram is thus extensive and flexible. An overexposed plate can be irradiated with infra-red till it reaches a readable intensity. Likewise, an underexposed plate can be increased in intensity to a readable density by increasing the infra-red radiation.

Figure 6 shows the storage sensitivity curve in the x-ray region of the spectrum for a particular phosphor. Each type of storage phosphor has its own sensitivity curve and exhibits its own characteristic properties as well. Although a blue-green phosphor was used extensively in these experiments, it is now possible to obtain a phosphor for any color region in the visible spectrum. Blue-green is a comfortable color to the eye, but it is not the most efficient color for definition.

In general, the results of these experiments have revealed numerous possibilities of infra-red phosphors in radiography. Their properties, however, are not always advantageous.

Referring back to Figs. 3 and 4, we recognize the same undesirable characteristics of graininess and diffusion as in the fluorescent screen. This might be corrected to a large extent by more refined processing methods. Figure 2 shows another disadvantage—a reluctant release of stored energy, because of the exponential decay rate, which may handicap the speed of repeated exposures upon a given phosphor plate. For all practical purposes, however, the energy remaining in the phosphor after reasonable exposure to infra-red radiation is negligible and the contrast of the recorded image appears to diminish with time.

Phosphorography will not immediately replace the x-ray film nor the fluorescent screen. It will supplement these methods and be particularly valuable in special cases. For example: flash fluoroscopy will be possible with the high current condenser-discharge x-ray tubes, whereby an image made with a very short x-ray exposure of high intensity will be made immediately visible, and retained long enough for inspection or recording on paper or film.

Grateful acknowledgment is extended to Drs. R. E. Ward and R. T. Ellickson of the Polytechnic Institute of Brooklyn for their willing cooperation in supplying information and material for this work.

# A Flux Plotting Method for Obtaining Fields Satisfying Maxwell's Equations, with Applications to the Magnetron

PRESCOTT D. CROUT
*Radiation Laboratory,* Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Received September 13, 1946)

In this report the flux plotting methods which have previously been applied to fields satisfying Laplace's and Poisson's equations are extended to fields satisfying Maxwell's equations. Application is made to the hole and slot, and vane types of magnetron tubes.

## 1. INTRODUCTION

"FLUX plotting" is a procedure for drawing two mutually orthogonal families of curves in such a way that the rectangles so formed have a specified geometric shape at each point. This shape is so determined that the curves represent a field which satisfies a specified partial differential equation. This equation, which characterizes the field, is not used directly, but it or its equivalent is used only indirectly to determine the geometrical condition that is used in actually constructing the plot. Flux plotting procedures for obtaining fields satisfying Laplace's equation, Poisson's equation, and certain other equations have been devised and applied successfully to wide varieties of problems.[1] The purpose of this paper is to extend these methods to fields characterized by Maxwell's equations. In so doing we do not use these equations directly, but use Ampere's law and Faraday's law instead. These laws together with Stoke's theorem evidently give Maxwell's equations. Some familiarity with flux plotting methods is assumed.

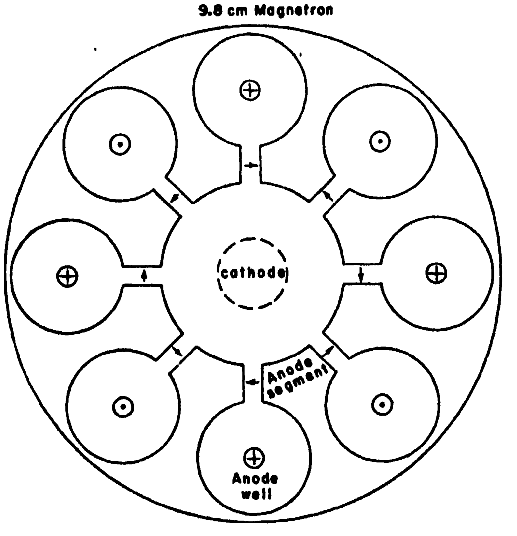


FIG. 1. Cross section of hole and slot type magnetron. Magnetic field is axial (perpendicular to the plane of the paper). Its directions in the anode wells at a certain instant are shown by the circles ⊕ and ⊙. Electric field is across the gaps between the anode segments. Its directions at a certain instant are shown by the arrows.

For the sake of clarity the new method will be described in connection with the specific application which gave rise to it—that of an infinite, hole and slot type magnetron, a cross section of which is shown in Fig. 1, operating in its main mode ($+ - + -$ or $\Pi$ mode).[2] Later in Section 6 application is made to the vane type magnetron. All fields are sinusoidally alternating. The electric current which links each anode well (cavity) and passes across the gap between two adjacent anode segments produces the axial magnetic field in the well (Ampere's law); also the rate of change of this axial magnetic field produces the voltage which maintains the electric current (Faraday's law). Because of symmetry we need consider only the field which lies between two axial planes which bisect, respectively, an anode segment and an anode well, Fig. 2.

---

* This paper is based on work done for the Office of Scientific Research and Development under contract OEMsr-262 with the Massachusetts Institute of Technology.

[1] H. Poritsky, "Graphical field-plotting methods in engineering," Trans. A.I.E.E. 57, 727 (1938). B. Hague, "Méthodes Analytiques, Graphiques et Expérimentales Utilisées Pour l'Étude des Champs Magnétiques et Électriques dans Les Machines et Appareils Électriques," Comptes rendus 4, 47 (1932). (This reference contains an extensive bibliography.) A. D. Moore, *Fundamentals of Electrical Design*, p. 31. P. D. Crout, "The Determination of Fields Satisfying Laplace's, Poisson's, and Associated Equations by Flux Plotting," Radiation Lab. Rep. 1047.

[2] The field so obtained can be made the basis of calculations for determining the frequency spectrums of finite strapped or unstrapped magnetrons. See Crout and Bothwell, "A Method for Calculating Magnetron Resonant Frequencies and Modes," Radiation Lab. Rep. 1039.

## 2. NATURE AND CONSTRUCTION OF THE FLUX PLOT

The peak electric field is shown in Fig. 2 by means of a number of electric lines of force, which divide the region into corresponding tubes of flux; and a number of lines of no work, which are orthogonal to the lines of force, and which divide the region into corresponding strips. The peak magnetic field is perpendicular to the plane of the paper. Let the various tubes and strips be numbered consecutively as indicated, and let

$L_{ij}$ = length of elementary rectangle common to $i$th tube and $j$th strip measured along a line of force through the center;
$W_{ij}$ = width of elementary rectangle common to $i$th tube and $j$th strip measured along a line of no work through the center;
$A_{ij}$ = area of elementary rectangle common to $i$th tube and $j$th strip;
$\sigma_i\phi$ = electric flux in the $i$th tube, $\phi$ being the arbitrarily chosen flux per unit tube; (1)
$E$ = electric intensity = electric flux density since dielectric constant = 1;
$H$ = magnetic intensity = magnetic flux density since permeability = 1;
$\omega$ = angular frequency;
$c$ = velocity of light.

### Gaussian units are used throughout

We shall first consider the equilibrium of an elementary rectangle, and shall leave till later the consideration of the elementary triangles which occur at the "kernel" or point of intersection of the lines of no work. The total peak displacement current flowing outside (on the cathode side) of the center of rectangle $ij$ is[3]

$$\left(\tfrac{1}{2}\sigma_i+\sum_{k=1}^{i-1}\sigma_k\right)\frac{\phi\omega}{4\pi};\tag{2}$$

hence, noting Ampere's law, we see that the value of $H$ at this point is independent of $j$, and is given by

$$\left(\tfrac{1}{2}\sigma_i+\sum_{k=1}^{i-1}\sigma_k\right)\frac{\phi\omega}{c}.\tag{3}$$

The total magnetic flux in the $j$th strip between the center of rectangle $ij$ and the kernel is therefore

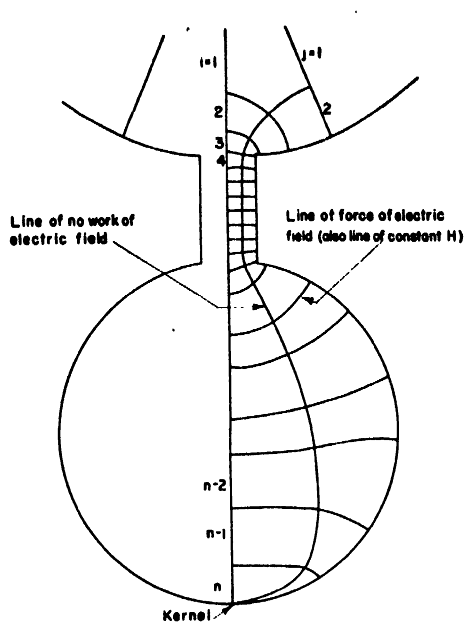

i=1
j=1
2
2
3
4
Line of no work of electric field
Line of force of electric field (also line of constant H)
n-2
n-1
n
Kernel

FIG. 2. Electric field in hole and slot type magnetron.

$$\frac{\phi\omega}{2c}\left(\tfrac{1}{2}\sigma_i+\sum_{k=1}^{i-1}\sigma_k\right)A_{ij}+\frac{\phi\omega}{c}\sum_{l=i+1}^{n-1}\left(\tfrac{1}{2}\sigma_l+\sum_{k=1}^{l-1}\sigma_k\right)A_{lj}+\frac{\phi\omega}{c}\left(\alpha_j\sigma_n+\sum_{k=1}^{n-1}\sigma_k\right)A_{nj},\tag{4}$$

where

$A_{ij}=L_{ij}W_{ij}$ for *rectangle ij*,
$L_{nj}$ = base of triangle in $j$th strip,
$W_{nj}$ = altitude of triangle in $j$th strip,
$A_{nj}$ = area of triangle in $j$th strip = $(L_{nj})_{\text{average}}W_{nj}$, (5)
$\alpha_j$ = fraction of $W_{nj}$ from the last line of force to the center of gravity of the last triangle (equals $\frac{1}{3}$ for triangles with straight sides).

By using the center of gravity as the point at which to evaluate $B$ in an elementary rectangle or triangle, multiplication of this value of $B$ by the elementary area gives the flux exactly if the surface which represents $B$ is a plane, regardless of whether or not that plane is horizontal ($B$ constant). The first term in (4) gives the peak magnetic flux in half of rectangle $ij$, whereas the last gives that in the triangle.

---

[3] From here on the word "peak" will be omitted, it being understood that all quantities are peak values of sinusoidally varying quantities. Complex notation can be used if desired, in which case the magnetic intensity will be purely imaginary.

Noting Faraday's law we see that the electric flux in the $i$th tube can be obtained from the rate of change of the magnetic flux given by (4), thus

$$\sigma_i\phi=\left\{\frac{\phi\omega}{c}\left[\tfrac{1}{2}\left(\tfrac{1}{2}\sigma_i+\sum_{k=1}^{i-1}\sigma_k\right)A_{ij}+\left(\alpha_j\sigma_n+\sum_{k=1}^{n-1}\sigma_k\right)A_{nj}+\sum_{l=i+1}^{n-1}\left(\tfrac{1}{2}\sigma_l+\sum_{k=1}^{l-1}\sigma_k\right)A_{lj}\right]\right\}\frac{\omega W_{ij}}{cL_{ij}}. \tag{6}$$

Here the brace is the magnetic flux given by (4), which with $\omega$ gives the rate of change of this flux, or the induced voltage in abvolts.[8] Division by $cL_{ij}$ gives the electric intensity in statvolts/cm, and multiplication by $W_{ij}$ gives the total electric flux in the rectangle, which must equal $\sigma_i\phi$. Cancelling $\phi$ and transposing, (6) becomes

$$\frac{L_{ij}\sigma_i}{W_{ij}}=\left(\frac{\omega}{c}\right)^2\left\{\tfrac{1}{2}\left(\tfrac{1}{2}\sigma_i+\sum_{k=1}^{i-1}\sigma_k\right)A_{ij}+\left(\alpha_j\sigma_n+\sum_{k=1}^{n-1}\sigma_k\right)A_{nj}+\sum_{l=i+1}^{n-1}\left(\tfrac{1}{2}\sigma_l+\sum_{k=1}^{l-1}\sigma_k\right)A_{lj}\right\}, \tag{7}$$

which is the desired equilibrium condition for the *rectangle*.

We shall next obtain the equilibrium condition for a triangular area at the kernel. Let the triangular area Fig. 3 be broken up into elementary bands of constant width $dx$ by a suitable family of curves, $x$ being measured along the orthogonal trajectory which passes through the center of gravity. We shall suppose that $H$ is essentially constant over the triangle, its value $H_0$ being taken as that at the center of gravity; also we shall suppose that the length of the elementary band at $x$ is given closely by

$$L=Kx^m, \tag{8}$$

where $K$ and $m$ are constants. The total magnetic flux in the triangle from the vertex to the elementary band is

$$\int_0^x H_0Kx^m dx=\frac{H_0Kx^{m+1}}{m+1};$$

hence the electric intensity at the band is

$$\left(\frac{1}{Kx^m}\right)\left(\frac{H_0Kx^{m+1}}{m+1}\right)\left(\frac{\omega}{c}\right), \tag{9}$$

in which the last two factors give the voltage across the band. Integrating (9) we obtain $\sigma_n\phi$, the total electric flux in the triangle, thus

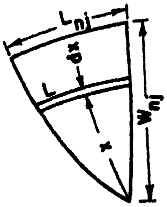


FIG. 3. Triangular area at kernel.

$$\sigma_n\phi=\int_0^{W_{nj}}\frac{H_0\omega x}{c(m+1)}dx=\frac{H_0\omega W_{nj}^2}{2c(m+1)}. \tag{10}$$

Noting the last term of (4) we replace $H_0$ by its value

$$\frac{\phi\omega}{c}\left(\alpha_j\sigma_n+\sum_{k=1}^{n-1}\sigma_k\right), \tag{11}$$

and obtain, after cancelling $\phi$ and transposing,

$$W_{nj}^2=\frac{2(m+1)\sigma_n}{\left(\frac{\omega}{c}\right)^2\left(\alpha_j\sigma_n+\sum_{k=1}^{n-1}\sigma_k\right)}. \tag{12}$$

Dividing the left hand side by $L_{nj}W_{nj}$, and the right by its equal $KW_{nj}^{m+1}$, and noting that

$$A_{nj}=KW_{nj}^{m+1}/(m+1),$$

we finally obtain

$$\frac{L_{nj}\sigma_n}{W_{nj}}=\frac{1}{2}\left(\frac{\omega}{c}\right)^2\left(\alpha_j\sigma_n+\sum_{k=1}^{n-1}\sigma_k\right)A_{nj}, \tag{13}$$

which is the desired equilibrium condition for the *triangle*. We note that this expression contains neither $K$ nor $m$; also, in comparing (13) with (7) we note the presence of the factor $\frac{1}{2}$ in (13).

As usual with flux plotting methods, the actual process of making the plot is one of systematic trial and error. The choice of the $\sigma_i$'s is a matter of convenience. If $\sigma_i$ is put equal to one, the $i$th tube is a unit tube; however, at the edge of the plot and in the weaker parts of the field it is desirable to use fractional tubes so that more lines show on the plot.

In constructing the plot a step-by-step process can be used whereby the values of $L_{ij}$, $W_{ij}$, and $\sigma_i$ taken from one approximate plot are placed in (7) and (13) to obtain improved values of $L_{ij}/W_{ij}$ and $\sigma_i$, which are then made the basis of the plot which constitutes the next approximation. The first plot would be largely assumed from physical considerations. One step of the step-by-step process will now be described.

First we note that the $\sigma_i$'s associated with a plot may be multiplied by any constant without altering the plot. This is evident from (7) and (13), and also from the fact that the effect is merely to magnify the field strength. Values of $L_{ij}$, $W_{ij}$, and $\sigma_i$ taken from the given approximate plot are now placed in the right hand sides of (7) and (13) to give for each rectangle or triangle the value of $L_{ij}\sigma_i/W_{ij}$ to be used in constructing the next approximate plot.[4] In so doing the unknown factor $(\omega/c)^2$ is replaced by an arbitrary constant multiplier $k$; hence the values obtained are of the form $k\beta_{ij}$ where the $\beta_{ij}$'s are computed numbers. Dividing by the measured values of $L_{ij}/W_{ij}$ we obtain a set of values of $k\beta_{ij}W_{ij}/L_{ij}$, which we denote by $\bar{\sigma}_{ij}$. If $k$ can be chosen so that $\bar{\sigma}_{ij}=\sigma_i$ for all rectangles and triangles, then this value of $k$ can be taken as $(\omega/c)^2$, all conditions are satisfied, and the given plot is final. Actually this cannot be done except with the final plot; hence in general we choose the new set of $\sigma_i$'s so that $\sigma_i$ approximates all the $\bar{\sigma}_{ij}$'s in the $i$th tube. Since $\sigma_i$ is of the form $k\gamma_i$ this is equivalent to choosing $\gamma_i$ so as to approximate all the values of $\beta_{ij}W_{ij}/L_{ij}$ in the $i$th tube. This can be done by choosing an average value or perhaps better, a weighted average with the areas as weighting factors. Dividing the computed values of $L_{ij}\sigma_i/W_{ij}$, namely $k\beta_{ij}$, by these new values of $\sigma_i$ we obtain the values $\beta_{ij}/\gamma_{ij}$, which we use as a new set of $L_{ij}/W_{ij}$ values in constructing the new plot.[4] We note that these are independent of $k$, which may be given any convenient numerical value.

The step-by-step process ends when a value of $k$ can be chosen for which $\bar{\sigma}_{ij}=\sigma_i$ for all rectangles and triangles, this value being $(\omega/c)^2$, as stated above.[5] The most accurate value of $\omega/c$ is probably given by equating the total electric energy to the total magnetic energy obtained from the final plot, as described in Section 4.[3]

### 3. ALTERNATIVE PROCEDURE USING A PLANIMETER

Let

$s_j$ = distance from the kernel measured along a line of no work at the center of the $j$th strip, (14)

$\Phi=\sigma\phi$ = total electric flux outside of the line of force at $s_j$. Here $\phi$ is again the arbitrarily chosen "normal" flux per tube.

$A_j$ = area of $j$th strip from the kernel to $s_j$;

then proceeding as with (3), noting Ampere's law, the magnetic flux density for any $s_j$ is

$$\Phi\omega/c=\sigma\phi\omega/c. \tag{15}$$

The total magnetic flux in $A_j$ is therefore

$$\frac{\omega\phi}{c}\int_0^{A_j}\sigma dA_j,$$

it being permissible to consider $\sigma$ a function of $A_j$ since both $\sigma$ and $A_j$ are functions of $s_j$. Continuing as with (6), noting Faraday's law, we multiply by $\omega/c$ to obtain the voltage

$$\phi\left(\frac{\omega}{c}\right)^2\int_0^{A_j}\sigma dA_j \tag{16}$$

across the strip at $s_j$. Choosing for $s_j$ the value at the center of the $i$th tube, dividing by $L_{ij}$ to obtain the electric intensity, and multiplying by $W_{ij}$ we obtain the flux $\sigma_i\phi$ of the tube, thus

$$\sigma_i\phi=\phi\left(\frac{\omega}{c}\right)^2\frac{W_{ij}}{L_{ij}}\int_0^{\bar{A}_{ij}}\sigma dA_j,$$

where $\bar{A}_{ij}$ is the strip area to the center of the $i$th tube. Cancelling $\phi$ and transposing we ob-

---

[4] Each computed value of $L_{ij}\sigma_i/W_{ij}$ is associated with the center of its rectangle or the center of gravity of its triangle, respectively. If the center of a new rectangle differs appreciably from that of the old, interpolation in this field of values may be necessary. The same is true of triangles.

[5] From the derivation of (7) and (13) or, later, of (17) and (18) it is evident that the above method of successive substitutions is equivalent to using the first approximate electric field to determine the corresponding magnetic field, corresponding to which, in turn, is obtained an electric field, which is taken as the second approximation—and so on. A method of successive substitutions may not be desirable in some cases, in which the new set of $L_{ij}/W_{ij}$ values may be chosen arbitrarily after comparing the new values computed as above with those of the preceding plots.

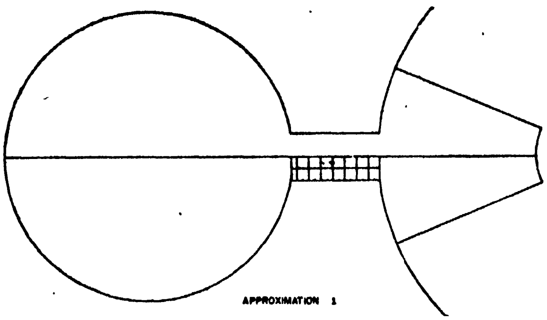


FIG. 4. Electric field in hole and slot type magnetron. Approximation 1.

tain finally

$$\frac{L_{ij}\sigma_i}{W_{ij}}=\left(\frac{\omega}{c}\right)^2\int_0^{A_{ij}}\sigma dA_j, \tag{17}$$

which is the desired equilibrium condition for a *rectangle.*

We shall now derive a condition analogous to (17) which can be applied to the triangles at the kernel. Denoting by $\mathcal{E}_j$ the voltage, given by (16), across the strip at the base of the $j$th triangle, and noting from (9) that for any approximating function $Kx^m$ for triangle width the field intensity varies linearly with $x$, we have for the total electric flux in the triangle

$$\sigma_n\phi=\frac{\mathcal{E}_j}{L_{nj}}\int_0^{W_{nj}}\frac{x}{W_{nj}}dx=\frac{\mathcal{E}_jW_{nj}}{2L_{nj}},$$

or, substituting from (16), cancelling $\phi$, and transposing,

$$\frac{L_{nj}\sigma_n}{W_{nj}}=\frac{1}{2}\left(\frac{\omega}{c}\right)^2\int_0^{A_{nj}}\sigma dA_j, \tag{18}$$

where the upper limit of the integral is the area of the triangle. This is the desired equilibrium condition for the *triangle.* In contrast to (17) we note the factor $\frac{1}{2}$ in (18), also the fact that $L_{nj}$ and $W_{nj}$ are the base and altitude of the triangle, respectively.

In applying (17) and (18) a convenient set of $\sigma_i$'s for the various tubes is chosen, and $A_j$ is obtained with a planimeter for values of $s_j$ corresponding to lines of force on the plot. These values of $A_j$ are then plotted against the corresponding known values of $\sigma$. This curve gives $\sigma$ as a function of $A_j$; hence a planimeter or approximate integration can be used to give the integrals in (17) and (18). Such a curve is drawn for each strip.

The step-by-step process described at the end of Section 2 evidently applies here, (17) and (18) being used instead of (7) and (13). As before $\omega/c$ is not determined until the final plot has been made, after which $\omega/c$ is obtained by equating the total electric and magnetic energies as described in Section 4.

## 4. ELECTRIC AND MAGNETIC FIELD ENERGY. DETERMINATION OF $\omega/c$

In any strip the electric flux passing through an elementary band bounded by two lines of force is $\phi d\sigma$. Dividing by $ds_j$ we obtain the electric intensity. The total electric energy in the strip is therefore[6]

Electric energy in $j$th strip

$$=\frac{\phi^2}{8\pi}\int_0^{A_j}\left(\frac{d\sigma}{ds_j}\right)^2dA_j. \tag{19}$$

Noting (15) we also have

Magnetic energy in $j$th strip

$$=\frac{\phi^2}{8\pi}\left(\frac{\omega}{c}\right)^2\int_0^{A_j}\sigma^2dA_j. \tag{20}$$

Using the method of Section 3 a curve of $\sigma$ as a function of $s_j$ can be plotted, and $d\sigma/ds_j$ obtained by approximate differentiation; also a curve of $\sigma$ as a function of $A_j$ is already available.[7] The integrals in (19) and (20) can hence be obtained

---

[6] The expression for electric energy obtained by means of the magnetic field using (16) merely duplicates (20), as is shown by the following.

Electric energy in $j$th strip

$$=\frac{1}{2}\int_0^{\sigma}\left[\phi\left(\frac{\omega}{c}\right)^2\int_0^{A_j}\sigma dA_j\right]\frac{\phi d\sigma}{4\pi}$$

$$=-\frac{\phi^2}{8\pi}\left(\frac{\omega}{c}\right)^2\int_0^{A_j}\left[\int_0^{A_j}\sigma dA_j\right]\frac{d\sigma}{dA_j}dA_j.$$

Integrating by parts noting that $\sigma=0$ when $A_j$ takes its greatest value, this becomes

Electric energy in $j$th strip

$$=-\frac{\phi^2}{8\pi}\left(\frac{\omega}{c}\right)^2\left\{\left[\sigma\int_0^{A_j}\sigma dA_j\right]_0^{A_j}-\int_0^{A_j}\sigma^2dA_j\right\}$$

$$=\frac{\phi^2}{8\pi}\left(\frac{\omega}{c}\right)^2\int_0^{A_j}\sigma^2dA_j,$$

which is (20).

[7] Tables for approximate differentiation are given in Crout, "An application of polynomial approximation to the solution of integral equations arising in physical problems," J. Math. and Phys. 19, 34 (1940).

using a planimeter or approximate integration. Adding the energies of the various strips the total electric energy and the total magnetic energy can be obtained, both being proportional to $\phi^2$. Equating these, $\phi^2$ cancels out, and we have

$$\frac{\omega}{c}=\left[\frac{\sum^{j}\int_0^{A_j}\left(\frac{d\sigma}{ds_j}\right)^2 dA_j}{\sum^{j}\int_0^{A_j}\sigma^2 dA_j}\right]^{\frac{1}{2}}, \qquad (21)$$

the upper limits on the integrals being the strip areas, and the summations being over the strips. This method of determining $\omega/c$ reminds one of Raleigh's method in elastic vibration problems, and offers probably the most accurate way of using the flux plot to obtain the resonance frequency.

We shall finally obtain expressions for the electric and magnetic energies which are applicable if the individual rectangles and triangles are considered separately, as in Section 2. The electric energy in a rectangle is evidently

Electric energy in rectangle

$$=\frac{1}{2}\left(\frac{\sigma_i\phi}{4\pi}\right)\left(\frac{\sigma_i\phi L_{ij}}{W_{ij}}\right)=\frac{\sigma_i^2\phi^2 L_{ij}}{8\pi W_{ij}}; \qquad (22)$$

also, noting (3), the magnetic energy of a rectangle is seen to be

Magnetic energy in rectangle

$$=\frac{\phi^2}{8\pi}\left(\frac{\omega}{c}\right)^2\left(\tfrac{1}{2}\sigma_i+\sum_{k=1}^{i-1}\sigma_k\right)^2 L_{ij}W_{ij}. \qquad (23)$$

From (9) it follows that in a triangle the electric intensity is of the form $K_1x$ where $K_1$ is a constant. $K_1$ can be determined by the relation

$$\sigma_n\phi=\int_0^{W_{nj}}K_1x\,dx=\frac{K_1W_{nj}^2}{2},$$

whence the electric intensity becomes

$$E_{nj}=\frac{2\sigma_n\phi x}{W_{nj}^2}.$$

The electric energy in the triangle is therefore, noting (8),

Electric energy in triangle

$$=\frac{1}{8\pi}\int_0^{W_{nj}}\left[\frac{2\sigma_n\phi x}{W_{nj}^2}\right]^2 Kx^m dx$$

$$=(\sigma_n\phi)^2 K W_{nj}^{m-1}/2\pi(m+3),$$

Electric energy in triangle $\qquad (24)$

$$=(\sigma_n\phi)^2 L_{nj}/2\pi(m+3)W_{nj},$$

where $L_{nj}$ and $W_{nj}$ are the base and altitude of the triangle, respectively, and $K$ and $m$ are the values applicable to the $j$th triangle.

In the triangle the magnetic field is closely equal to the constant value (11); hence the magnetic energy is closely

Magnetic energy in triangle

$$=\frac{\phi^2}{8\pi}\left(\frac{\omega}{c}\right)^2\left(\alpha_j\sigma_n+\sum_{k=1}^{n-1}\sigma_k\right)^2\int_0^{W_{nj}}Kx^m dx$$

Magnetic energy in triangle $\qquad (25)$

$$=\frac{\phi^2}{8\pi}\left(\frac{\omega}{c}\right)^2\left(\alpha_j\sigma_n+\sum_{k=1}^{n-1}\sigma_k\right)^2\frac{L_{nj}W_{nj}}{(m+1)}.$$

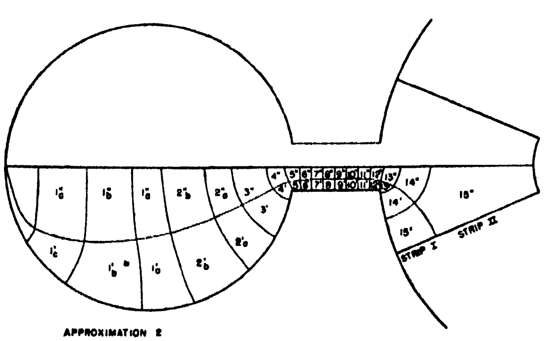


FIG. 5. Electric field in hole and slot type magnetron. Approximation 2.

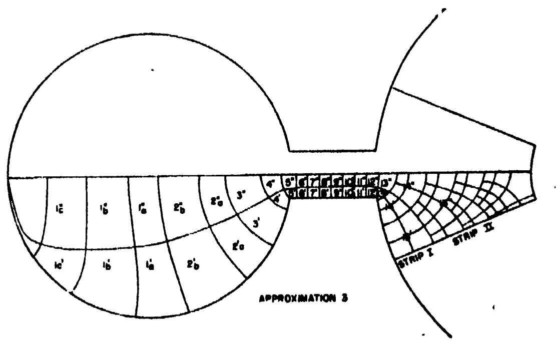


FIG. 6. Electric field in hole and slot type magnetron. Approximation 3.

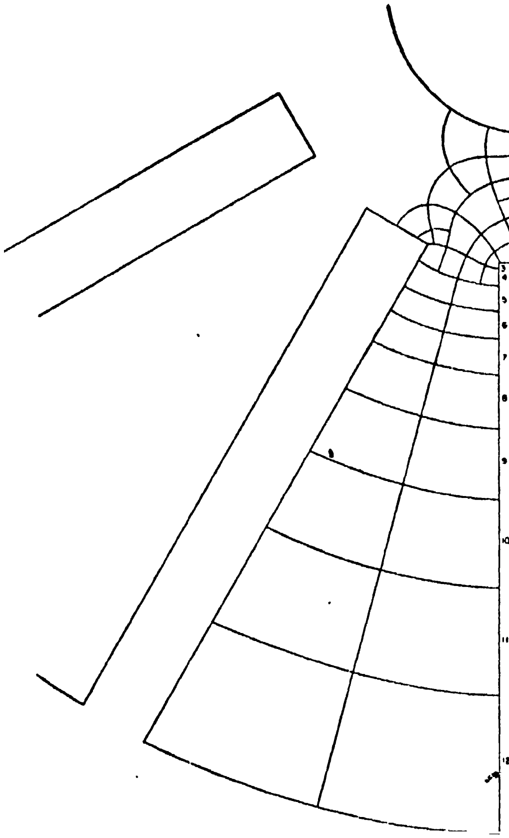


FIG. 7. Electric field in vane type magnetron.

### 5. APPLICATION TO A SPECIFIC PROBLEM. HOLE AND SLOT TYPE MAGNETRON

The procedure used to obtain the field in the cavity shown in Fig. 2 will now be outlined.[8] The approximation that was used for the electric field in order to get the step-by-step process started was a uniform electric field in the gap between two adjacent anode segments, the electric field being taken as zero everywhere else, Fig. 4. Data from this first approximation are now inserted in the right-hand sides of (17) and (18) as described in Section 3 to obtain corrected data for the plot which constitutes the second approximation. In so doing we note that $\sigma$ is zero in the anode-cathode region, and increases linearly through the gap to a value which remains constant in the well. It follows that $\int_0^{\bar{A}_{ij}}\sigma dA_j$ and hence the computed values of $L_{ij}\sigma_i/W_{ij}$ are constant in the anode-cathode region, and that in the well $L_{ij}\sigma_i/W_{ij}$ and $L_{nj}\sigma_n/W_{nj}$ are proportional, respectively, to $\bar{A}_{ij}$ and $A_{nj}/2$. The flux plot which constitutes the second approximation was begun at the inner (cathode) side of the gap, the work being carried both inward and outward from this point. We note that in the anode-cathode region this plot can be a lattice of curvilinear squares, which gives a field satisfying Laplace's equation since $L_{ij}\sigma_i/W_{ij}$ is constant in this region; also we note that the choice of the $\sigma_i$'s is a matter of convenience. The second approximation is shown in Fig. 5.[9]

Continuing the step-by-step process, data from this second approximation are now inserted in (17) and (18) to obtain data for a third approximation. In so doing each new value of $\sigma_i$ is taken as the average of the $\bar{\sigma}_{ij}$'s of the elementary areas in the $i$th tube. After the third approximation, Fig. 6, was obtained, it was found that the ratio of $\sigma_i$ to $W_{ij}\beta_{ij}/L_{ij}$ was constant for all the elementary areas to within the accuracy expected of the graphical process. This third approximation, Fig. 6, was therefore taken as the final result, the constant ratio being $(\omega/c)^2$. The frequency corresponding to the value so obtained for $\omega/c$ is $2.818\cdot 10^9$ c.p.s.; that obtained by equating electric and magnetic energies using (21) was $2.825\cdot 10^9$ c.p.s. The various tube dimensions were

| | |
|---|---|
| Cathode radius | 0.1182″ |
| Anode radius | 0.3175″ |
| Gap width | 0.066″ |
| Radius to center of well | 0.640″ |
| Well diameter | 0.398″. |

The values of the $\sigma_i$'s for the final plot, Fig. 6, are shown in Table I.

### 6. APPLICATION TO A SPECIFIC PROBLEM. VANE TYPE MAGNETRON

If the shape of the electric field is known approximately, the work required to obtain a plot can be greatly reduced.[10] For example in the

---

[8] The work outlined in this section including that involved in constructing the flux plots shown in Figs. 5 and 6 was done by Mr. F. S. Holt.

[9] The labels shown in Figs. 5 and 6 are those that were used in making the plot. The strips are indicated by primes, the tubes are numbered, and subdivisions are indicated by letters or, if obvious, are not indicated. This notation differs from that used in the derivation of (7), (13), (17), and (18); but is convenient for actual construction.

[10] The work outlined in this section including that involved in constructing the flux plot Fig. 7 was done by Mr. F. E. Bothwell.

TABLE I. Values of $\sigma_i$.

| Tube | $\sigma_i$ | Tube | $\sigma_i$ | Tube | $\sigma_i$ |
|---|---|---|---|---|---|
| Triangle | .0625 | 4 | .999 | 11 | .999 |
| 1c | .125 | 5 | .999 | 12 | 1.00 |
| 1b | .250 | 6 | .999 | 13 | 1.00 |
| 1a | .250 | 7 | .985 | 14 | 1.00 |
| 2b | .499 | 8 | .989 | 15 | .836 |
| 2a | .499 | 9 | .994 | | |
| 3 | .997 | 10 | .997 | | |

TABLE II.

| | Values of $\sigma_i$ | | | | |
|---|---|---|---|---|---|
| step / $i$ | 1 | 2 | 3 | 4 | 5 |
| 1 | 1.00000 | .97989 | .97990 | .97998 | .97999 |
| 2 | 1.00000 | .97444 | .97484 | .97489 | .97491 |
| 3 | 1.00000 | .96696 | .96786 | .96792 | .96793 |
| 4 | 1.00000 | .96162 | .96293 | .96302 | .96300 |
| 5 | 1.00000 | 1.14617 | 1.14824 | 1.14833 | 1.14835 |
| 6 | 1.00000 | 1.00887 | 1.01120 | 1.01130 | 1.01127 |
| 7 | 1.00000 | 1.06811 | 1.07140 | 1.07148 | 1.07150 |
| 8 | 1.00000 | 1.24744 | 1.25118 | 1.25127 | 1.25128 |
| 9 | 1.00000 | 1.32741 | 1.33002 | 1.33001 | 1.33002 |
| 10 | 1.00000 | 1.14551 | 1.14404 | 1.14388 | 1.14386 |
| 11 | 1.00000 | .83281 | .82418 | .82390 | .82388 |
| 12 | 1.00000 | .34074 | .33420 | .33401 | .33400 |

Table of values of $k$

| step | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $k$ | | 6.0374 | 5.7562 | 5.7517 | 5.7516 |

vane type magnetron, Fig. 7, the electric lines of force in most of the region between the vanes evidently approximate arcs of circles centered at the cathode; also in the anode cathode region, where the magnetic field is weak, the electric field approximates a solution of Laplace's equation. With these facts in mind Fig. 7 was drawn for the case where the cathode and all vanes but one are at the same voltage, no attempt being made to satisfy (7). Omitting the line of no work let us temporarily consider the field as occupying a single strip; then Eqs. (7) form a set of twelve homogeneous, linear, algebraic equations for the $\sigma_i$'s. These can be conveniently solved by successive substitutions by first placing all $\sigma_i$'s on the right-hand side of (7) equal to unity. The $\sigma_i$'s so obtained are inserted in the right-hand sides of (7) to give the next approximation, and so on. In each step the $k$ which replaces $(\omega/c)^2$ was given that value which would make the average of the computed $\sigma_i$'s equal to unity. This method of successive substitutions has the physical significance described in footnote 5; its convergence was rapid, and the results obtained are shown in Table II. When put equal to $(\omega/c)^2$ the last obtained value of $k$ corresponds to a wave-length of 2.6200 cm. The (probably) more exact value obtained by equating total electric and magnetic energies as described in Section 4 is 2.6456 cm.

Let us finally try to improve the accuracy of the plot by subdividing and proceeding as in Section 5. We therefore draw the line of no work that was omitted above; and, placing the $\sigma_i$'s of Step 5 in the right-hand side of (7), determine a new set of values $L_{ij}/W_{ij}$. These new values, however, duplicate the old values in all tubes except the first four, in which there is a discrepancy of one percent. Since $L_{ij}$ and $W_{ij}$ cannot be measured to within one percent, no alteration of the plot is indicated, and it is therefore taken as final. It is probable that the procedure used in this section on the vane type magnetron can be used to advantage on the hole and slot type magnetron considered in Section 5.

# A Magnetron Oscillator for Dielectric Heating

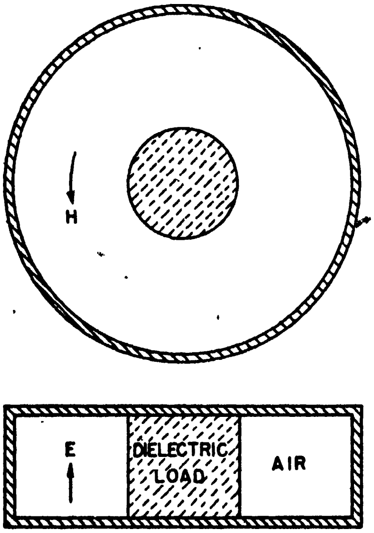
R. B. Nelson
*Research Laboratory, General Electric Company, Schenectady, New York*

(Received January 31, 1947)

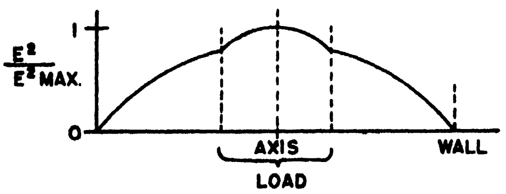
Magnetron oscillators can provide very high frequency power for dielectric heating. Factors determining the choice of frequency for particular jobs are discussed. The design of a magnetron having 5 kw continuous output at 1050 megacycles is described, and performance data on the tube are given.

## INTRODUCTION

THE method of heating non-conducting materials by dielectric loss in a high frequency electric field is becoming an important industrial process. Many of the materials heated by this method, such as plastics, are of such low loss that it is hard to heat them as rapidly as desired. The rate of heating increases with the electric field strength and, in practically all materials, with the frequency. Since the usable field is definitely limited by arcing between the electrodes, the only remaining way to speed up the heating process is to increase the frequency.



GEOMETRY OF CAVITY AND LOAD



RADIAL VARIATION OF FIELD

FIG. 1. Distribution of heating in cylindrical load. In a simple cylindrical cavity where the electric field is produced by radial currents in the walls, the r.m.s. field is a maximum at the center and decreases radially as a Bessel function. The rate of decrease depends upon the wave-length in the dielectric, giving the discontinuity in slope of the curve of $E^2$ plotted against radial distance. Rate of production of heat per unit volume is proportional to $E^2$.

Commercial dielectric heaters have been restricted to frequencies below about 40 megacycles by the limitations of available vacuum tubes. However, the recent development of magnetron oscillators has made available considerable power at frequencies up to some 10,000 megacycles. Using magnetrons for heating thus allows the choice of almost any desired frequency.

## CHOICE OF FREQUENCY

It must not be assumed that the highest frequencies are always best. While heating rates go up with frequency, a point is reached where non-uniformity of heating limits the usable frequency. Two effects contribute to non-uniformity. First, in the case of high loss materials such as most foods, there is the decreased heating toward the center of large objects because the power is absorbed by the surface layers before it can penetrate. Second, in the case of low loss materials such as plastics, the wave-length of the heating power limits the volume of material that may be heated uniformly. To apply power efficiently to low loss material, it is necessary to surround the material with reflecting walls so that the electromagnetic wave passing through it is not radiated and lost after one passage. The reflecting walls set up a standing wave in and around the load material. This wave has in general a three-dimensional space variation of electric field, with periodic maxima and minima.

Since the dielectric heating at any point is proportional to the square of the electric field strength, widely varying temperatures will be produced in the material if its dimensions are comparable with a wave-length. As an example of how the standing-wave pattern limits the usable frequency, consider a cylinder of load material located axially in a cylindrical resonant cavity, as shown in Fig. 1. Exciting the cavity in its lowest mode with axial electric field, the field is a maximum at the center and falls off radially as

$$E = E_{\max} J_0(6.28r/\lambda),$$

where $E$ is the electric field strength at radius $r$ from the axis, $E_{\max}$ is the field on the axis, and $\lambda$ is the wave-length in the material. If it is required to heat the edges of the load 90 percent as much as the center,

$$E^2/E_{\max}^2 = J_0^2(6.28r/\lambda) \geqq .90 \text{ at the edges,}$$

and solving the above equation for a load of radius $a$

$$a \leqq .0725\lambda.$$

If, for example, the load is a piece of phenolic plastic with dielectric constant 4.0 and radius 1.0 cm,

$$\lambda = \lambda_0/4.0^{\frac{1}{2}},$$

where $\lambda_0$ is the free space wave-length corresponding to the frequency used. Solving for $\lambda_0$

$$\lambda_0 \geqq 27.6 \text{ cm}.$$

Thus, the maximum frequency allowed in this case would be 1087 megacycles.

### OSCILLATOR DESIGN CONSIDERATIONS

On the basis of considerations such as the above example, it appeared that many dielectric heating processes could be done well at a frequency of about 1000 mc. Therefore, a magnetron oscillator was developed in the General Electric Research Laboratory to give 5 kw output at 1050 mc. In designing this oscillator, consideration had to be given to several special requirements of heating apparatus which differ from other uses of magnetrons.

The most important specific requirement is that the heating oscillator must be capable of operating into any load impedance, including the completely unloaded condition. This is caused by the non-uniformity of successive charges of load material, to the certainty of bad adjustment of matching transformers by unskilled operators, and the possibility of accidentally running the oscillator with no load at all. To meet these conditions, the anode should be capable of dissipating the entire input power to the tube. The output seal and transmission line to the load must stand very high voltages and currents associated with high standing wave ratios. The cathode should not be damaged by the excessive power it receives from back-bombardment by electrons, which is a function of the load impedance as well as the plate voltage and current.

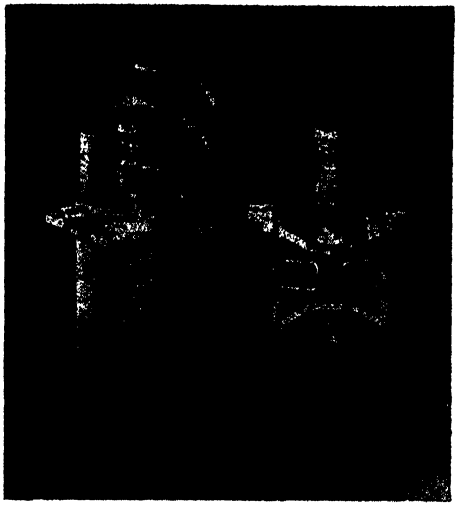


FIG. 2. Photograph of the 5-kw magnetron, showing the external appearance and also an axial section.

Although the multiple-cavity magnetron has many resonant modes, efficient oscillation occurs only in the "$\pi$" mode where each anode is 180° out of phase with those adjoining it. When this mode is overloaded, as may occur by poor load matching, the oscillator will tend to jump to another mode, which is usually completely unloaded, and may continue in the spurious mode, even if the mismatch is corrected, until it is turned off and restarted. The accepted cure for mode jumping is to strap the anodes together in two alternate sets with low impedance conductors. This raises the frequency of all modes except the "$\pi$" mode to a point where the oscillator does not readily excite them. The

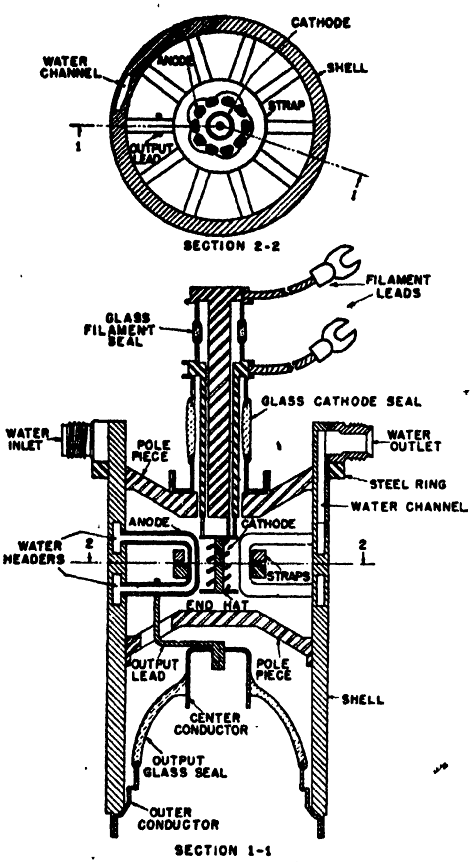


FIG. 3. Construction of water cooled magnetron. Sectional views showing the tubular anodes and water circulating channels. The O.D. of the copper shell is 3½″.

heating magnetron, as distinguished from the radar transmitter, runs continuously, so once it has jumped modes it does not have a chance to restart properly. Also the wide range of load impedances favors mode jumps. These factors necessitate very tight strapping. It is desirable that the next higher frequency mode be at least twice the frequency of the operating mode. In the oscillator to be described, this separation was attained by a new method of strapping.

### A 5-KW, 1050-MC MAGNETRON

The magnetron is shown in the photograph, Fig. 2, and constructional details in Fig. 3. It has 10 anodes, and so may be run at relatively low voltage and magnetic field. Rated operating conditions for the full 5 kw output are: plate voltage 5.0 kv, plate current 1.5 amp., magnetic field 1500 gauss. Water cooling was adopted for the anodes, and this allowed making the anodes of U-shaped pieces of copper tubing through which the water flows. Channels in the copper tube shell form headers for water distribution.

In addition to providing excellent cooling, the tubular anodes have a minimum of inter-anode capacity, thus allowing a large fraction of the capacitance of the resonant circuits to be provided by low impedance straps. The resulting tight strapping gives a frequency ratio of 1:2 between the operating mode and the next higher mode. The double ring straps are looped inside the anode U's. This allows the straps to be attached near the centers of the anode faces, eliminating unbalanced r-f voltages on the anodes. Also, the anodes shield the cathode and the rest of the tube from the fundamentally unbalanced double ring straps. The ten resonant circuits formed by the anode loops extending radially inward from the tube shell are electrically similar to conventional vane resonators.

The over-all size of the magnetron is determined by the output seal, which must be large to

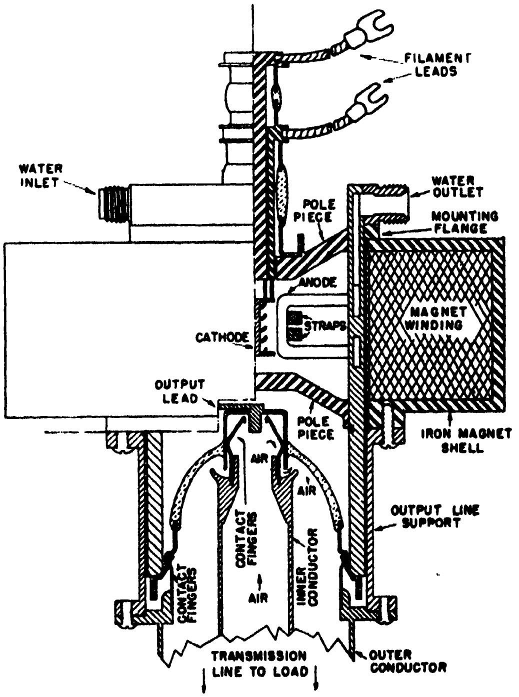


FIG. 4. Magnetron mounting and connections. The tube is slid downward into the cylindrical magnet, simultaneously plugging in to the coaxial output transmission line which is rigidly mounted on the magnet. The iron magnetic circuit is shown in heavy cross-hatching.

handle the power. In this cylindrical structure the seal occupies the entire diameter of the tube shell. The coaxial arrangement of the seal with the cathode and magnetic pole pieces allows the tube to be plugged into its magnet and output transmission line simultaneously. Figure 4 shows the method of mounting the tube. An electromagnet is used, consisting of a solenoidal coil with an iron shell whose ends line up with the iron pole pieces inside the tube. Air is blown through the center conductor of the output line to cool the seal, the inverted cup shape of the seal giving a good distribution of the cooling air as well as protecting the seal from breakage.

The power output is obtained by attaching a flat strap from the center conductor of the seal to one of the anodes. The exact point of attachment is selected to provide proper loading of the oscillator when the 52.5-ohm transmission line on the output has no standing wave.

The cathode is a simple spiral of pure tungsten wire with molybdenum end hats. It requires 600 watts heating power. In a magnetron having considerable electron bombardment of the cathode, it is desirable to use a very inefficient high wattage cathode, so that the back heating becomes a small fraction of the total heating power.

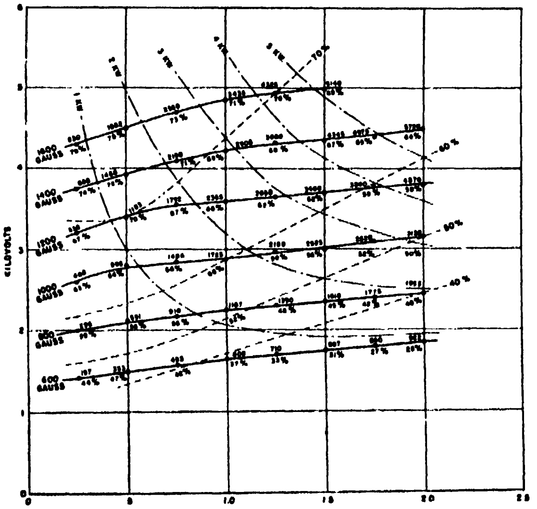


FIG. 5. Performance chart. Solid lines are contours of constant magnetic field, dashed lines are constant power, and dotted lines constant efficiency. Power, in watts, and efficiency measured are indicated on the experimental points. This chart is taken with a load matched to the transmission line.

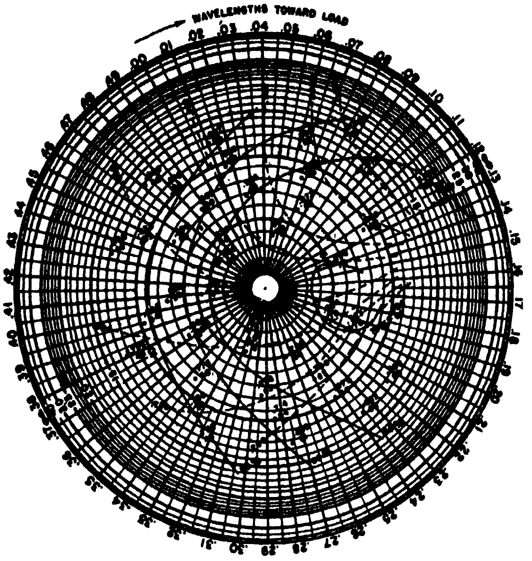


FIG. 6. Rieke diagram. The effect of load impedance on the efficiency and wave-length of the oscillator is best shown by this chart. The data were taken at fixed conditions of 1.0 amp. plate current and 1200 gauss magnetic field. Upper figures are wave-length (add 29.000 cm) and lower figures percentage efficiency. Constant efficiency and constant wave-length contours are shown by solid and dotted lines, respectively. The radial coordinate is voltage reflection coefficient, with the scale marked in standing wave ratio in decibels. The azimuthal coordinate is the phase angle of the reflection coefficient, with the scale marked in wave-lengths displacement of the standing wave minimum from an arbitrary zero.

## MAGNETRON OPERATING CHARACTERISTICS

The performance of the magnetron as an oscillator is graphically presented by several charts. Shown in Fig. 5 is the "performance chart" in which power output (upper numbers) and efficiency (lower numbers) are plotted against plate voltage and current. Lines of constant magnetic field are also shown.

Another important oscillator characteristic is shown in Fig. 6, the "Rieke diagram," in which efficiency and wave-length are plotted as a function of load impedance for constant current and magnetic field. The radial coordinate is standing wave ratio in the output line (in voltage) plotted on a scale proportional to reflection coefficient. The angular coordinate is the phase angle of the standing wave minimum. Thus the position of a point on the chart represents a complex load impedance, which might be transformed into resistance and reactance components by superposition of a Smith chart.

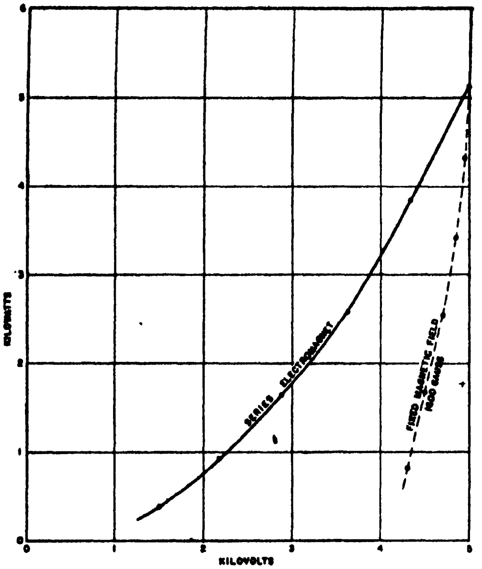


FIG. 7. Effect of plate voltage on output power. With a constant magnetic field as obtained from a permanent magnet, power output varies rapidly with plate voltage, following the dotted curve. An electromagnet excited by the magnetron plate current reduces this variation.

It is apparent from Fig. 5 that the plate current changes very rapidly with plate voltage at constant magnetic field. That is, the variational resistance for plate modulation is around 250 ohms. This is a good situation if one desires to use amplitude modulation, as the modulating power is low. However, in dielectric heating, a constant r-f output is desired, and account must be taken of variations in line voltage. It is obvious from the curves that a small variation in line voltage and hence in plate voltage will produce a large change in plate current and in output power. Figure 7 shows this relation, neglecting the regulation of the plate supply rectifier.

A method of greatly reducing the effect of line voltage fluctuations and, incidentally, plate supply ripple, is to run the tube's electromagnet in series with the low voltage end of the plate supply. This gives a magnetic field directly proportional to the plate current. Since the plate voltage is almost directly proportional to the magnetic field, the plate current will in turn be proportional to plate voltage. The resultant ohmic plate resistance characteristic produces the variation of r-f power with plate supply voltage shown by the solid curve in Fig. 7.

The relationship between d.c. plate current, plate voltage, and magnetic field also affects the way in which the magnetron output depends on the load impedance. The data plotted in Fig. 6 were taken under the easily reproducible conditions of constant current and magnetic field, letting the voltage find its own value. The voltage varies much less at constant current than the current would at constant voltage.

In practical operation we do not have either a constant current or constant voltage power supply, and the regulation characteristics of the supply affect the relationship of magnetron power output and load impedance. It is obvious that this is a very important relationship for heating applications.

With the magnetic field supplied by a series electromagnet, some of the effects of power supply regulation are simulated, and r-f power output tends to be stabilized. For a constant voltage power supply and series magnet, the variation of output power with load for a pure resistive load is plotted in Fig. 8. For comparison, there is also shown the relations obtained with constant voltage power supply, and with constant current.

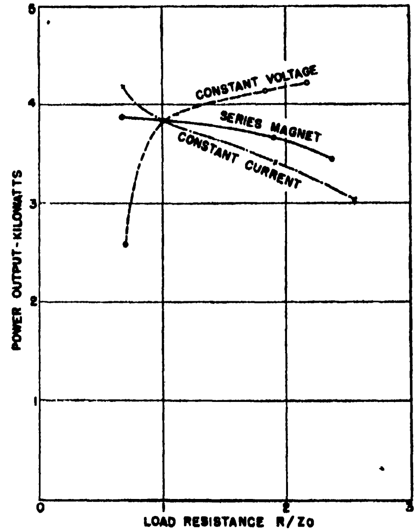


FIG. 8. Effect on output power of varying pure resistance loads. For loads which present a pure resistive impedance to the magnetron, the power delivered is plotted as a function of the ratio of load resistance $R$ to the characteristic impedance $Z_0$ of the transmission line. At constant magnetic field, the power increases with load resistance if the magnetron plate voltage is held constant, but decreases if the current is constant. Using a series electromagnet with constant voltage supply, the better characteristic shown by the solid line is obtained.

It has been found that if plate voltage is applied to the magnetron with no magnetic field but with the magnet in series, the tube will start in a very high frequency mode of oscillation, drawing plate current so that the voltage never rises enough to start the desired mode of circuit oscillation. To prevent this, it is necessary to have a field of at least 600 gauss present before plate voltage is applied. A rectifier across the magnet as shown in Fig. 9 supplies this starting current, and normally stops supplying current when the magnetron current builds up a voltage drop across the magnet coil greater than the output voltage of the rectifier.

The magnet supply rectifier can be made to do another useful job as a fine control for plate current and output power. In the plate supply rectifier, continuous voltage control is expensive, but a few taps for step control are cheap. Now at any given plate voltage, the magnetron is approximately a constant-magnetic-field device, independent of plate current. If the plate current supplies part of the magnetic field and the rest comes from the magnet rectifier, any increase in rectifier current will tend to increase the magnetic field and thus will cause a corresponding decrease in magnetron current. In this way continuous adjustment of output power is obtained by regulating the 100-watt magnet rectifier. Figure 9 shows the Variac connection used for control.

## R-F LOADING EQUIPMENT

The problem of carrying the 1050 mc power from the magnetron and coupling it into a load is one which has almost as many answers as there are varieties of material to be heated. The power is taken from the oscillator on a stub-supported coaxial line, the start of which is shown in Fig. 4. A line size of 3⅛" O.D. with 1¼" center conductor has been adopted as adequate for this power level and frequency. For most loads, a large step-up in impedance is necessary to match the line to the load. This usually takes the form of a resonant cavity tuned to the operating frequency and coupled to the transmission line by a loop or antenna.

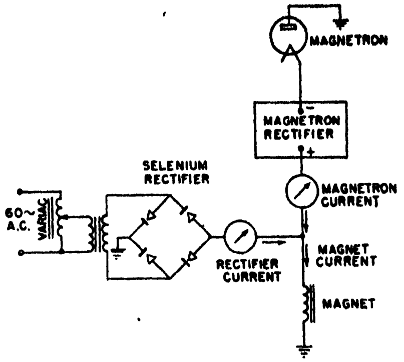


FIG. 9. Current supply for electromagnet. An auxiliary d.c. supply for the magnet is necessary to assure proper starting. This variable-voltage rectifier may be used to control the magnetron power output.

## ACKNOWLEDGMENTS

The author wishes to thank Dr. A. W. Hull for his continued interest and advice in carrying out this development. Mr. T. C. Swartz, Miss H. C. Hertha, and Mr. P. J. Scavullo have contributed largely to its progress.

## REFERENCES

So much of the literature on modern magnetron developments is unpublished that no attempt has been made at references to original sources. For a general discussion of magnetron principles and a pre-war bibliography, refer to J. B. Fisk, H. D. Hagstrum, and P. L. Hartman, "The magnetron as a generator of centimeter waves," Bell Sys. Tech. J. **25**, 167 (April 1946).

# An Approximate Solution of the Differential Equation of the Ultracentrifuge

W. J. ARCHIBALD
*Dalhousie University, Halifax, Nova Scotia*

(Received November 13, 1946)

In determining molecular weights with the ultracentrifuge by the equilibrium method, it is necessary to drive the rotor at relatively high speeds for very long periods of time. It would be a real advantage to be able to deduce molecular weights from the distribution of concentration within a cell placed in a rotor without waiting for equilibrium. The solution of a certain differential equation gives the theoretical distribution within the cell at any time, and a comparison of the experimentally determined concentrations with those deduced from theory should yield molecular weights without waiting for equilibrium. The exact solution of this differential equation is known, but it is not suitable for numerical work. The present paper presents an approximate solution from which numerical results can be obtained with ease.

ONE of the most important methods of measuring high molecular weights is by the use of the ultracentrifuge as developed by Svedberg and Beams. A dilute solution of the substance whose molecular weight is to be determined is placed in a small cell, with radial sides and cylindrical ends, near the periphery of the rotor, and under the influence of the centrifugal force the distribution of concentration changes from an initial uniform distribution to an exponential one. When this final equilibrium state has been reached the concentration "$c$," at any point in the cell, is given by the expression

$$c = c_0 \exp\left[\frac{M(1-V\rho)\omega^2(r^2-r_0^2)}{2RT}\right], \tag{1}$$

where "$M$" is the molecular weight of the dissolved substance, "$V$" its partial specific volume, $\rho$ the density of the solvent, $\omega$ the angular velocity of rotation, and "$r$" the distance of any point in the cell from the axis of rotation. "$C_0$" is the concentration at the point "$r_0$." "$R$" is the gas constant per mole and "$T$" the absolute temperature. The equilibrium state is reached when the tendency for solute particles to settle to the outer part of the cell because of the centrifugal force is exactly equal to the tendency for them to diffuse inwards because of the concentration gradient. It is obvious that Eq. (1) gives "$M$" in terms of quantities that can be measured, although the experiment is one that can only be performed successfully with faultless optical and mechanical equipment.

The time required for the attainment of the equilibrium state is often very long—sometimes the centrifuge must run for many days, or even weeks, at constant speed and at a constant temperature. If the time of a run could be shortened this method of molecular weight determination would be much more useful. The experimental difficulties would be greatly reduced and the number of runs made with any centrifuge would be increased. This latter consideration is an important one in view of the many substances that could be studied by this method. If the distribution of concentration within the cell at any time after the start of an experiment could be deduced theoretically, a comparison of the measured and computed distribution might enable "$M$" to be found without waiting for equilibrium. The function $c(r, t)$, which gives the concentration at any point at any time, would be required for this purpose and this is obtained by solving the following differential equation:

$$\frac{1}{r}\frac{\partial}{\partial r}\left\{\left(D\frac{\partial c}{\partial r}-\omega^2 rsc\right)r\right\}=\frac{\partial c}{\partial t}. \tag{2}$$

In (2), "$D$" is the diffusion constant of the solute and "$s$" the sedimentation constant, i.e., the velocity of settling of the dissolved substance in unit gravitational field. The quantities "$D$" and "$s$" are related to each other. It is possible to show that

$$D/s = RT/M(1-V\rho). \tag{3}$$

A proof of this relation is given in a previous paper by the author.[1] The differential Eq. (2) was

---

[1] W. J. Archibald, Phys. Rev. 53, 746 (1938).

first derived by Lamm.[2] A simple derivation is to be found in reference 1.

A treatment of Eq. (2) which is free of any mathematical approximations has been given and is to be found in two places in the literature.[3,4] While the solution given is exact and can be employed for purposes of computation, yet the cost in time of getting numerical results is so great that little use has been found for it. The expression obtained for $c(r,t)$ is simple in form and would be most serviceable if the mathematical functions which it contains could be looked up in tables. However such tables are not available and the task of compiling them would be a major project in applied mathematics. It is the purpose of this paper to derive another expression for $c(r,t)$ which, while not exact, is felt to be good enough for many purposes, and which contains functions that are tabulated.

In dealing with Eq. (2) it is more convenient to use "$z$" and "$\tau$" rather than "$r$" and "$t$" as the independent variables, where

$$z=\frac{\omega^2 s}{2D}r^2=\frac{M(1-V\rho)\omega^2}{2RT}r^2 \quad \text{and} \quad \tau=2\omega^2 st. \tag{4}$$

As $r_1 \leq r \leq r_2$ where $r_1$ and $r_2$ are the inner and outer limits of the cell, respectively, then $a \leq z \leq b$ where

$$\begin{Bmatrix} a \\ b \end{Bmatrix}=\frac{M(1-V\rho)\omega^2}{2RT}\begin{Bmatrix} r_1^2 \\ r_2^2 \end{Bmatrix}.$$

In terms of these variables we find

$$\frac{c(z,\tau)}{c_0}=\frac{b-a}{e^b-e^a}e^z+\sum_{n=1}^{\infty}\frac{\int_a^b e^{-z}M(\alpha_n,1,z)dz}{\int_a^b e^{-z}\{M(\alpha_n,1,z)\}^2dz} \times M(\alpha_n,1,z)e^{(\alpha_n-1)\tau}, \tag{5}$$

where $M(\alpha,\gamma,z)$ is a solution of the differential equation

$$\frac{d^2M}{dz^2}+\left(\frac{\gamma}{z}-1\right)\frac{dM}{dz}-\frac{\alpha}{z}M=0 \tag{6}$$

and the $\alpha_n$'s are the "eigenvalues" of the parameter $\alpha$. The $\alpha_n$'s are those particular values of $\alpha$ which give the functions $M(\alpha_n,1,z)$ the property that

$$\frac{d}{dz}M(\alpha_n,1,z)=M(\alpha_n,1,z) \tag{7}$$

for $z=a$ and $z=b$. "$c_0$" is the initial uniform concentration. (The reader is asked to consult reference 3 or 4 for a detailed proof of (5) and a full explanation of the results stated here.) $M(\alpha_n,1,z)$ is an abbreviation for

$$F(\alpha_n,1,z)-\frac{F(\alpha_n,2,a)}{W(\alpha_n,2,a)}W(\alpha_n,1,z), \tag{8}$$

where $F(\alpha_n,\gamma,z)$ and $W(\alpha_n,\gamma,z)$ are the two independent solutions of (6) which are to be found written out in full.[3,4] It is the difficulty of obtaining numerical values for the $F$ and $W$ functions which limits the usefulness of Eq. (5).

It will be our purpose to see if it is possible to derive some expression for the functions $M(\alpha_n,1,z)$, other than that given in (8), and which contains only well-known functions. This alternative expression will of necessity be an approximation and a careful study of the magnitude of the different quantities involved will be required to make the approximation a good one. A further purpose will be to discover a simple method for determining the eigenvalues $\alpha_n$.

In (5) the first term, in $e^z$, gives $c/c_0$ for the equilibrium state. In an actual experiment it is not desirable to have the final concentration vary too widely throughout the cell. If one were to accept as a reasonable set of conditions that finally the concentration at $r_2$ be three times that at $r_1$ then

$$e^b/e^a \simeq 3, \quad \text{or} \quad (b-a)\simeq 1.1.$$

This serves to show the narrow limits within which the variable "$z$" is confined. Furthermore, since

$$b/a=r_1^2/r_2^2,$$

we have

$$a=(b-a)\frac{r_1^2}{r_2^2-r_1^2}\simeq 1.1\frac{r_1^2}{r_2^2-r_1^2}.$$

The cells used are usually small and one might

---

[2] Lamm, Ark. Mat. Astron. Fysik. 21B, No. 2 (1929).
[3] W. J. Archibald, Phys. Rev. 54, 371 (1938).
[4] W. J. Archibald, Ann. N. Y. Acad. Sc. 43, 211 (1942).

choose as typical values the following: $r_2=5.0$ cm and $r_1=4.5$ cm. This gives $a\simeq 4.7$. As a result of these considerations we are entitled to treat the variable "$z$" as being moderately large (of the order of 5.0) and as being confined within a short range of values. The results of a computation on a typical experiment[3,4] show that $5.0\leq z\leq 6.2$ and these figures are probably typical of most cases in which the equilibrium method is used.

A further point of interest has to do with the quantities $\alpha_n$. The exact theory gives $\alpha=1$ as one of the eigenvalues. This $\alpha$, of course, gives rise to the first term in (5). All the other $\alpha$'s have large negative values. In the computations referred to above it was found that $\alpha_1=-38.7$. $\alpha_2$ was not determined, but the indications were that it was less than $-100$.

The first term in (5) presents no difficulty: but the terms in the summation part of the expression are bothersome. Thus we will confine our attention to these terms. This means that we must study Eq. (6) (with $\gamma=1$), remembering that $\alpha$ is very large and negative and that "$z$" is moderately large and of restricted variation. In (6) let

$$M=e^{z/2}z^{-\frac{1}{2}}V. \tag{9}$$

The equation for "$V$" then becomes

$$\frac{d^2V}{dz^2}+\left\{-\frac{1}{4}+\frac{1}{4z^2}-\frac{(\alpha-\frac{1}{2})}{z}\right\}V=0. \tag{10}$$

The considerations given above suggest that it is permissible to drop the term $1/4z^2$. This term is very small in comparison with $-(\alpha-\frac{1}{2})/z$ and is even quite small in comparison with $\frac{1}{4}$. Then (10) becomes

$$\frac{d^2V}{dz^2}+\left\{-\frac{1}{4}-\frac{(\alpha-\frac{1}{2})}{z}\right\}V=0.$$

This is still an inconvenient equation to deal with but it can be further simplified by replacing the term in braces by

$$\left\{-\frac{1}{4}\frac{(a+b)}{2}-(\alpha-\tfrac{1}{2})\right\}\frac{1}{z}=\frac{\sigma^2}{z}. \tag{11}$$

This term is a very close approximation to $-\frac{1}{4}-(\alpha-\frac{1}{2})/z$ when $z\simeq 5$, $\alpha$ is large and $b-a$ is of the order of unity. As an example we may use the numerical data from reference 3 or 4. Putting $\alpha=-38.7$, $a=5$, and $b=6.2$ with $5.0\leq z\leq 6.2$ we obtain the results given in Table I. It will be observed that the maximum difference between the two quantities is less than 0.5 percent.

TABLE I.

| $z$ | $f(z)$ | $f(z)$ approx. |
|---|---|---|
| 5.0 | 7.59 | 7.56 |
| 5.2 | 7.29 | 7.27 |
| 5.4 | 7.01 | 7.00 |
| 5.6 | 6.75 | 6.75 |
| 5.8 | 6.51 | 6.52 |
| 6.0 | 6.28 | 6.30 |
| 6.2 | 6.07 | 6.10 |

$$f(z)=-\tfrac{1}{4}-\frac{(\alpha-\frac{1}{2})}{z}$$

$$f(z)\text{ approx.}=\left\{-\tfrac{1}{4}\frac{(a+b)}{2}-(\alpha-\tfrac{1}{2})\right\}\frac{1}{z}$$

With $\alpha=-38.7$, $a=5.0$, and $b=6.2$.

Thus if one uses

$$\frac{d^2V}{dz^2}+\frac{\sigma^2}{z}V=0, \tag{12}$$

as the differential equation for the function "$V$" it seems reasonable to suppose that (9) will give a close representation of the function "$M$" for "$z$" between "$a$" and "$b$."

Now Eq. (12) has for its most general solution

$$V(z)=z^{\frac{1}{2}}\{AJ_1(2\sigma z^{\frac{1}{2}})+BN_1(2\sigma z^{\frac{1}{2}})\}, \tag{13}$$

where $J_1$ and $N_1$ are Bessel functions of the first and second kind, respectively, and $A$ and $B$ are arbitrary constants to be determined by the conditions of the problem. Combining (9) and (13) we have

$$M(\alpha,z)=e^{z/2}\{AJ_1(2\sigma z^{\frac{1}{2}})+BN_1(2\sigma z^{\frac{1}{2}})\}. \tag{14}$$

It must be emphasized that this expression can only be made to approximate $M(\alpha, 1, z)$ (by the proper choice of $A$ and $B$) for large values of $\alpha$ and for $z$ between $a$ and $b$. For any other range of variables it will undoubtedly be wide of the mark. Even subject to these restrictions there is no reason to suppose that $e^{z/2}J_1(2\sigma z^{\frac{1}{2}})$ is an approximation to $F(\alpha, 1, z)$ or $e^{z/2}N_1(2\sigma z^{\frac{1}{2}})$ to $W(\alpha, 1, z)$. All we are entitled to expect is that a linear combination of $F(\alpha, 1, z)$ and $W(\alpha, 1, z)$ can be replaced by some linear combination of the functions $e^{z/2}J_1(2\sigma z^{\frac{1}{2}})$ and $e^{z/2}N_1(2\sigma z^{\frac{1}{2}})$.

We will now adopt for the function $M(\alpha_n, 1, z)$, which occurs in Eq. (5), the form given in (14). However it is still necessary to discover what values of $\alpha_n$ to use. (It must be remembered that $\alpha$ is contained in the quantity $\sigma$.) It will be recalled that the function $M(\alpha, z)$ must be such that

$$dM/dz = M \tag{15}$$

for $z=a$ and $z=b$. This equation will be found to hold only when the $\alpha$'s are properly chosen. Let

$$Z_1(2\sigma z^{\frac{1}{2}}) = AJ_1(2\sigma z^{\frac{1}{2}}) + BN_1(2\sigma z^{\frac{1}{2}})$$

and in order to be specific let $N_1(2\sigma z^{\frac{1}{2}})$ be the Neumann function whose properties are well known. Then

$$M(\alpha, z) = e^{z/2} Z_1(2\sigma z^{\frac{1}{2}}). \tag{16}$$

Substituting (16) in (15) one gets

$$\frac{dZ_1}{d(2\sigma z^{\frac{1}{2}})} - \frac{z^{\frac{1}{2}}}{2\sigma} Z_1 = 0. \tag{17}$$

(In the following the argument of $Z$, $J$, and $N$ will be omitted.) But

$$\frac{dZ_1}{d(2\sigma z^{\frac{1}{2}})} = Z_0 - \frac{1}{2\sigma z^{\frac{1}{2}}} Z_1.$$

Therefore, (17) becomes

$$Z_0 - Z_1 \tan\zeta = 0, \tag{18}$$

where

$$\tan\zeta = \frac{1}{2\sigma}\left(\frac{1}{z^{\frac{1}{2}}} + z^{\frac{1}{2}}\right). \tag{19}$$

In terms of the $J$ and $N$ functions (18) becomes

$$A(J_0 - J_1 \tan\zeta) + B(N_0 - N_1 \tan\zeta) = 0. \tag{20}$$

At this point it is useful to make a further approximation. We have carried the Bessel and Neumann functions as far as it is convenient—to carry them further leads to complicated expressions. However the argument $2\sigma z^{\frac{1}{2}}$ is so large that they may be replaced by their asymptotic representations. The quantity $2\sigma z^{\frac{1}{2}}$ will usually have values in excess of 20 and in this range we may use the following:

$$\left.\begin{aligned} J_0 &= (\pi\sigma z^{\frac{1}{2}})^{-\frac{1}{2}} \cos(2\sigma z^{\frac{1}{2}} - \pi/4), \\ N_0 &= (\pi\sigma z^{\frac{1}{2}})^{-\frac{1}{2}} \sin(2\sigma z^{\frac{1}{2}} - \pi/4), \\ J_1 &= (\pi\sigma z^{\frac{1}{2}})^{-\frac{1}{2}} \cos(2\sigma z^{\frac{1}{2}} - 3\pi/4), \\ N_1 &= (\pi\sigma z^{\frac{1}{2}})^{-\frac{1}{2}} \sin(2\sigma z^{\frac{1}{2}} - 3\pi/4). \end{aligned}\right\} \tag{21}$$

Substituting these expressions in (20), and simplifying, one obtains

$$A(\cos\epsilon - \tan\zeta \sin\epsilon) + B(\sin\epsilon + \tan\zeta \cos\epsilon) = 0,$$

where $\epsilon = 2\sigma z^{\frac{1}{2}} - \pi/4$. Dividing through by $\cos\epsilon$ this becomes

$$A(1 - \tan\zeta \tan\epsilon) + B(\tan\zeta + \tan\epsilon) = 0$$

or

$$A + B\tan(\zeta + \epsilon) = 0. \tag{22}$$

Now $\zeta$ and $\epsilon$ are both functions of $z$ and equation (22) must hold for $z=a$ and $z=b$. Therefore, we get the two equations

$$\left.\begin{aligned} A + B\tan\{\zeta(a) + \epsilon(a)\} &= 0 \\ A + B\tan\{\zeta(b) + \epsilon(b)\} &= 0 \end{aligned}\right\}. \tag{23}$$

These equations can only have roots other than the trivial ones $A = B = 0$ if

$$\tan\{\zeta(a) + \epsilon(a)\} = \tan\{\zeta(b) + \epsilon(b)\}. \tag{24}$$

Equation (24) can be satisfied if

$$\{\zeta(b) + \epsilon(b)\} - \{\zeta(a) + \epsilon(a)\} = n\pi, \quad n = 0, 1, 2, \cdots \tag{25}$$

Introducing the functions $\zeta$ and $\epsilon$ explicitly (25) becomes

$$x = n\pi + \tan^{-1}\left\{\frac{(b^{\frac{1}{2}} - a^{\frac{1}{2}})}{x}\left(\frac{1+a}{a^{\frac{1}{2}}}\right)\right\} - \tan^{-1}\left\{\frac{(b^{\frac{1}{2}} - a^{\frac{1}{2}})}{x}\left(\frac{1+b}{b^{\frac{1}{2}}}\right)\right\}. \tag{26}$$

where $x = 2\sigma(b^{\frac{1}{2}} - a^{\frac{1}{2}})$. The permissible values of $\alpha$ are determined by finding the values of $x$ for which this equation is satisfied ($x$ is a function of $\alpha$). Before looking at (26) in detail it is obvious that when $n=0$ its solution is $x=0$. This means $\sigma=0$ or

$$\alpha = \tfrac{1}{2} - \tfrac{1}{8}(a+b).$$

This will give a small negative value for $\alpha$. Remembering that out treatment cannot be expected to be correct when $\alpha$ has a small absolute value it will be realized that this is the approximation to the eigenvalue $\alpha=1$. However the term in (5) arising from this eigenvalue (*viz.*, the term in $e^z$) is simple and there is no point in doing otherwise than taking it as it stands.

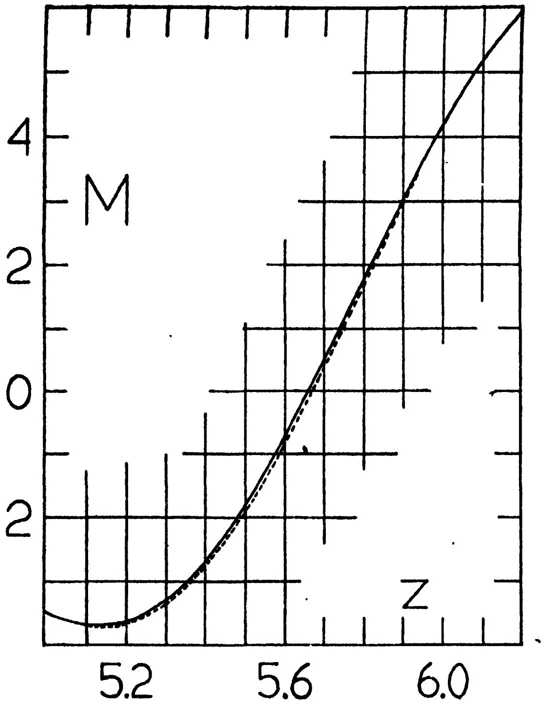


Fig. 1. The full line is the function $M(-38.7, 1, z)$ and the dotted line the function $-0.4205M(-38.77, z)$.

Since $b-a\simeq 1$, we have

$$(b^{\frac{1}{2}}-a^{\frac{1}{2}})\simeq\frac{1}{b^{\frac{1}{2}}+a^{\frac{1}{2}}}\simeq\frac{1}{2b^{\frac{1}{2}}}.$$

Therefore,

$$\frac{(b^{\frac{1}{2}}-a^{\frac{1}{2}})}{x}\frac{(1+b)}{b^{\frac{1}{2}}}\simeq\left(\frac{0.5}{x}\right)\frac{1+b}{b}\simeq\frac{0.5}{x}.$$

When $n=1, 2, \cdots$ in (26) an actual computation using representative figures shows that $x\simeq n\pi$. Thus the arguments of the inverse tan functions will be small, and these terms will be very small in comparison with $n\pi$. This being the case we will replace $x$ by $n\pi$ in the inverse tan functions and expand them retaining only the first term of the expansions. This gives

$$x=n\pi+\frac{(b^{\frac{1}{2}}-a^{\frac{1}{2}})}{n\pi}\left\{\frac{1+a}{a^{\frac{1}{2}}}-\frac{1+b}{b^{\frac{1}{2}}}\right\}$$

or

$$\sigma_n=\frac{n\pi}{2(b^{\frac{1}{2}}-a^{\frac{1}{2}})}+\frac{1}{2n\pi}\left\{\frac{1+a}{a^{\frac{1}{2}}}-\frac{1+b}{b^{\frac{1}{2}}}\right\}. \qquad (27)$$

It is interesting to see how accurately this formula gives $\alpha_1$ (i.e., $n=1$) for the problem referred to above in which $a=5.0$ and $b=6.2$. A simple computation gives $\sigma_1=6.154$ or $\alpha_1=-38.77$. The correct value is $\alpha_1=-38.70$. Thus it is seen that for large absolute values of $\alpha$ the approximation is a very good one. Using the exact theory, several weeks of computing were required to get $\alpha_1$ and the labor involved in getting $\alpha_2$ seemed prohibitive. Formula (27) gives a result, correct to within 0.2 percent in this case, with only a few minutes required for the computation. When $n=2$, we have $\sigma_2=12.36$ and $\alpha_2=-154.6$. This checks with the exact theory as to order of magnitude and is probably quite accurate because for such large values of $\alpha$ the approximations are especially valid. This eigenvalue and the ones which follow it are so large that rapid convergence of the summation in (5) is assured even for small values of $\tau$.

Returning to a consideration of Eq. (14) we see that it should now be written

$$M(\alpha_n, z)=A_n e^{z/2}\{J_1(2\sigma_n z^{\frac{1}{2}})+K_n N_1(2\sigma_n z^{\frac{1}{2}})\}, \qquad (28)$$

where

$$K_n=B_n/A_n=-\cot\{\zeta_n(a)+\epsilon_n(a)\}. \qquad (29)$$

Since it is not usual to find the $J_1$ and $N_1$ functions tabulated for such large values of the argument as occur in this problem it seems best to replace them by their asymptotic representations in $M(\alpha_n, z)$. Then (28) becomes

$$M(\alpha_n, z)=e^{z/2}z^{-\frac{1}{4}}\{\cos(2\sigma_n z^{\frac{1}{2}}-3\pi/4)$$
$$+K_n\sin(2\sigma_n z^{\frac{1}{2}}-3\pi/4)\}. \qquad (30)$$

It will be noted that the $A_n$ as well as some other constants which occur as multiplying factors have been dropped. That this is permissible is at once evident from Eq. (5). $M(\alpha_n, 1, z)$, or its approximation $M(\alpha_n, z)$, is found twice in the numerator and twice in the denominator and any constant multiplying factor will cancel. Hence $M(\alpha_n, z)$ as written in (30) will replace $M(\alpha_n, 1, z)$ as given by the exact theory, and the $\alpha_n$'s will be given by (27) and (11).

A noteworthy feature of the function $M(\alpha_n, z)$ is its simplicity and the ease with which numerical results can be obtained from it. Its validity is best illustrated by comparing the curve obtained from it with that obtained from the exact function $M(\alpha_n, 1, z)$. This latter function is tabulated for $5.0 \leq z \leq 6.2$ and $\alpha = -38.7$.[4] In Fig. (1) the two functions $M(-38.7, 1, z)$ and $-0.4205M(-38.77, z)$ are plotted for purposes of comparison. (It is to be observed that the two $M$ functions are not identical—one is a multiple of the other.) No attempt has been made to find the best multiplying factor: the factor $-0.4205$ is the one that makes the two functions identical at $z = 5.0$, the extreme end of the range. The usefulness of the approximation seems well attested within the range of values for which it was devised.

The evaluation of the integrals which occur in Eq. (5) does not seem of importance. Since $M(\alpha_n, z)$ must be computed in any case the simplest procedure is to evaluate these integrals by numerical integration.

Two sets of values of $c/c_0$ for several values of $\tau$ are to be found in Table II. In the first column of each group are the results given by the approximate functions and in the second column those given by the exact functions. These latter values are obtained from reference 4, for the problem considered there. In computing the values of $c/c_0$ from the exact solution the terms in $\alpha_2, \alpha_3, \cdots$ etc., were not used. This did not seem to introduce a serious error into the results because it was felt that these higher terms were insignificant even for such a small value of $\tau$ as .015. However, no proof of this could be given and it is reassuring to discover from the approximate solution that this omission is justified. If the term in $\alpha_2$ is included the numerical values most seriously affected are only changed by a few units in the third figure beyond the decimal. (It should be stated that the values given in Table II for the approximate solution do not include the terms in $\alpha_2, \alpha_3, \cdots$. They were omitted purposely in order to have the same conditions in both sets of calculations.) The greatest differences between the two sets of figures occur in the section headed $\tau = .015$. (This corresponds to a time equal to about 1/7 of that required to reach full equilibrium.) But even here the approximation is within 1.2 percent of being correct over most of the range of $z$. For larger values of $\tau$ the differences are much less.

TABLE II. $c(z, \tau)/c_0$.

| | $\tau = .015$ | | $\tau = .03$ | | $\tau = .05$ | | $\tau = .10$ | |
|---|---|---|---|---|---|---|---|---|
| $z$ | *A* | *B* | *A* | *B* | *A* | *B* | *A* | *B* |
| 5.0 | .726 | .717 | .632 | .627 | .569 | .567 | .524 | .524 |
| 5.1 | .795 | .784 | .695 | .689 | .627 | .625 | | |
| 5.2 | .853 | .841 | .754 | .747 | .687 | .684 | .640 | .639 |
| 5.3 | .900 | .889 | .809 | .804 | .748 | .746 | | |
| 5.4 | .938 | .926 | .863 | .857 | .813 | .810 | .778 | .777 |
| 5.5 | .968 | .956 | .916 | .910 | .882 | .879 | | |
| 5.6 | .994 | .983 | .971 | .965 | .955 | .953 | .944 | .944 |
| 5.7 | 1.021 | 1.013 | 1.030 | 1.026 | 1.037 | 1.035 | | |
| 5.8 | 1.051 | 1.050 | 1.096 | 1.095 | 1.126 | 1.126 | 1.148 | 1.148 |
| 5.9 | 1.094 | 1.099 | 1.174 | 1.177 | 1.228 | 1.229 | | |
| 6.0 | 1.157 | 1.165 | 1.269 | 1.273 | 1.344 | 1.346 | 1.398 | 1.398 |
| 6.1 | 1.241 | 1.254 | 1.382 | 1.389 | 1.476 | 1.479 | | |
| 6.2 | 1.358 | 1.374 | 1.519 | 1.528 | 1.628 | 1.632 | 1.705 | 1.705 |

Columns *A* contain the approximate values and columns *B* the exact values of the function $c(z, \tau)/c_0$.

To obtain a set of curves from the exact solution would take an experienced computer several weeks: using the functions derived in this paper the same curves could be obtained in six or eight hours.

# The Motion of a Conical Coil Spring

I. EPSTEIN*
*Department of Physics, University of Cincinnati, Cincinnati, Ohio*

(Received November 22, 1946)

The elongations and natural frequencies of conically-shaped coil springs are derived theoretically for several dynamic boundary conditions. Frequencies are verified experimentally on nickel and piano wire springs, attesting the reliability of the computations.

## 1. EQUATION OF MOTION

THE dissimilarities in the analytic expressions for the oscillatory motions of conically-shaped springs on one hand and cylindrical ones on the other, arise as results of the variability of mass per unit length in the direction of motion and of the inconstancy of the spring "constant" in the case of the conical type. If a constant mass per unit length of wire, $m$, is assumed, the total mass of a single complete turn of wire can be stated as a function of the diameter of the turn, or as a function of the diameter of the largest turn and the linear distance from it. Let it be required to find the mass of a complete turn of wire at an average distance $x$ from a coil of diameter $D_0$ (Fig. 1). The circumference of the turn in question is $\pi(D_0-2ax)$, where $a$ is the tangent of the half-angle at the vertex of the cone. Hence, its mass is $m$ times this value. The "inertial force" (mass times acceleration) of a single turn of wire is then, $\pi(D_0-2ax)m\partial^2E/\partial t^2$, where $E$ is the displacement of the turn (elongation of the spring), measured from its unstretched position. The inertial force of an element $dx$ must be equated to the net force necessary to give the element displacement $E$. The formula for the deflection of a single turn of wire (as measured by one 360° revolution of the coil around the longitudinal axis of the cone) of loop diameter $D$ and wire thickness $d$, under a load of $P$ pounds, is $f=8D^3P/Gd^4$, where $f$ is the deflection in inches, $D$ and $d$ are measured in inches, and $G$ is the torsional modulus of elasticity of the wire used.[1] Since the spring constant $K$ is defined as the force per unit deflection, the conical spring "constant" is $K=Gd^4/8D^3$, and the force on an element of wire due to spring tension is:

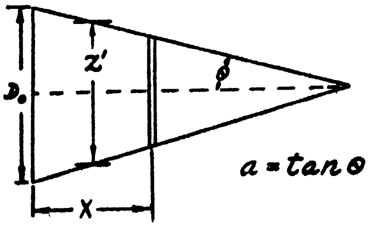


FIG. 1. Schematic diagram of spring (longitudinal section).

$$F=\frac{\partial}{\partial x}\left[K\frac{\partial E}{\partial x}\right]dx=\frac{\partial}{\partial x}\left[\frac{Gd^4}{8D^3}\frac{\partial E}{\partial x}\right]dx.$$

It is advantageous to use as independent variable the diameter of a turn of wire, defined by the equation $Z=D=D_0-2ax$. Thus, equating the inertial force to the tension force, the equation takes the form:

$$\frac{2\pi mZ^4}{a^2Gd^4}\frac{\partial^2E}{\partial t^2}=\frac{\partial^2E}{\partial Z^2}-\frac{3}{Z}\frac{\partial E}{\partial Z}. \tag{1}$$

A solution of the form $E(Z,t)=w(Z)\cdot T(t)$ is assumed, and substitutions are made back into Eq. (1). The variables $Z$ and $t$ may be separated and two ordinary differential equations arise. For the equation in $t$ we get:

$$d^2T/dt^2+\lambda^2T=0. \tag{2}$$

The general solution of this equation is $T=T_0\cos(\lambda t+\delta)$, where $\lambda=2\pi f$, where $f_j$ are the natural frequencies of the oscillatory system to be determined by boundary conditions operating on the ordinary differential equation in $Z$. For the equation in $Z$ we get:

$$\frac{-2\pi m\lambda^2}{a^2Gd^4}=\frac{1}{Z^4w}\frac{d^2w}{dZ^2}-\frac{3}{Z^5w}\frac{dw}{dZ}. \tag{3}$$

---

* Graduate student.

[1] *National Metals Handbook* (American Society for Steel Training, Cleveland, 1933), p. 990.

For ease in computation to be done later, the following changes are made at this point. $Z$ is changed so that $z=Z/D_0$. Thus, the range of the independent variable, rather than being from 0 to $D_0$, will now be from 0 to 1. Further, let $a^2Gd^4/2\pi mD_0^6$ be denoted by a single bulk constant $c$. Equation (3) then reduces to:

$$\frac{d^2w}{dz^2}-\frac{3}{z}\frac{dw}{dz}+\frac{\lambda^2z^4w}{c}=0. \tag{4}$$

Series solutions of Eq. (4) may be identified with solutions in Bessel functions of orders $\frac{2}{3}$ and $-\frac{2}{3}$ of the form:

$$W(y)=(y/2)^{\frac{2}{3}}\{AJ_{-2/3}(y)+BJ_{2/3}(y)\}, \tag{5}$$

where $y=\lambda z^3/3\sqrt{c}$. Finally, the general solution of Eq. (1), in terms of the independent variable $z$, is:

$$E(z,t)=(\Lambda/2)^{\frac{2}{3}}z^2\{AJ_{-2/3}(\Lambda z^3)+BJ_{2/3}(\Lambda z^3)\}\cos(\lambda t+\delta), \tag{6}$$

where $\lambda/3\sqrt{c}$ has been replaced by $\Lambda$. An investigation will now be conducted into the particular solutions of this equation to be used to satisfy specific boundary conditions.

## 2. SPECIAL BOUNDARY CONDITIONS

### I. Both Extremities Fixed

The particular solutions of Eq. (6) which satisfy the boundary conditions $E(0,t)=E(1,t)=0$, are:

$$E(z,t)=(\Lambda/2)^{\frac{2}{3}}z^2BJ_{2/3}(\Lambda z_3)\cos(\lambda t+\delta). \tag{7}$$

By taking $\delta=0$, we assume that the motion starts from rest, that is, $\partial E/\partial t=0$ at $t=0$. In order to satisfy the boundary condition at the base, it is necessary that $E(1,t)$ in Eq. (7) be zero for all $t$. This can be accomplished by limiting the values of $\Lambda$ to those which are roots of the equation $J_{2/3}(\Lambda)=0$. The first two roots of this equation are $\Lambda_1=3.38$ and $\Lambda_2=6.53$, while other roots may be deduced by adding $\pi$ successively to each former root.[2] Since $\Lambda_j=\lambda_j/3c^{\frac{1}{2}}$, and $\lambda_j=2\pi f_j$, then $f_j$ (the natural frequencies of the system under assigned boundary conditions) will be $f_j=3c^{\frac{1}{2}}\Lambda_j/2\pi$, where the $\Lambda_j$ are roots of $J_{2/3}(\Lambda)=0$. The constant $c$ may be determined in one of three ways. First, by the defining formula for $c$: $c=a^2Gd^4/2\pi mD_0^6$, using quantities which are deemed reliable for $a$, $G$, $d$, and $m$. Another method is to analyze the static case for an experimental determination of $Gd^4$ and using values for the other quantities to the best of one's knowledge. A third method is to solve for $\sqrt{c}$ using one experimentally determined natural frequency. Once determined in this manner, $c$ will retain the same value regardless of boundary conditions or initial distortions. It is best to arrive at a suitable value of $c$ by using all three methods simultaneously, and arriving at consistent values for all of the constants included in $c$.

TABLE I.

| | Root | $\Lambda_j$ | $f_j$(c/sec.) | $f_j$(experimentally) |
|---|---|---|---|---|
| Ni | 1st | 3.38 | 2.59 | 2.58 |
| | 2nd | 6.53 | 5.21 | |
| Piano wire | 1st | 3.38 | 2.53 | 2.55 |
| | 2nd | 6.53 | 5.08 | |

In experiments performed in conjunction with this problem, two springs (one of nickel and one of piano wire) were employed. By the methods just outlined, the value of $c$ adopted for the nickel coil was 2.75 and for the piano wire coil, 2.65. Thus, for the nickel coil,

$$f_j=\frac{3(2.75)^{\frac{1}{2}}\Lambda_j}{6.28}=\frac{\Lambda_j}{1.26},$$

and for the piano wire coil,

$$f_j=\frac{3(2.65)^{\frac{1}{2}}\Lambda_j}{6.28}=\frac{\Lambda_j}{1.29}.$$

Table I shows the results of this case.

### IIA. Vertex Fixed, Base Free

In this case, the boundary conditions to be met are that $E(0,t)=0$, and $\partial E/\partial x$ be zero at $z=1$. As in the previous case, in order to fulfill the condition at the vertex, $A$, (Eq. (6)), must be made identically zero. Thus, for a particular solution, we choose:

$$E(z,t)=(\Lambda/2)^{\frac{2}{3}}z^2BJ_{2/3}(\Lambda z^3)\cos(\lambda t). \tag{8}$$

---

[2] Dinnik, Archiv d. Math. und Phys. [3], 18, 337–338 (1911).

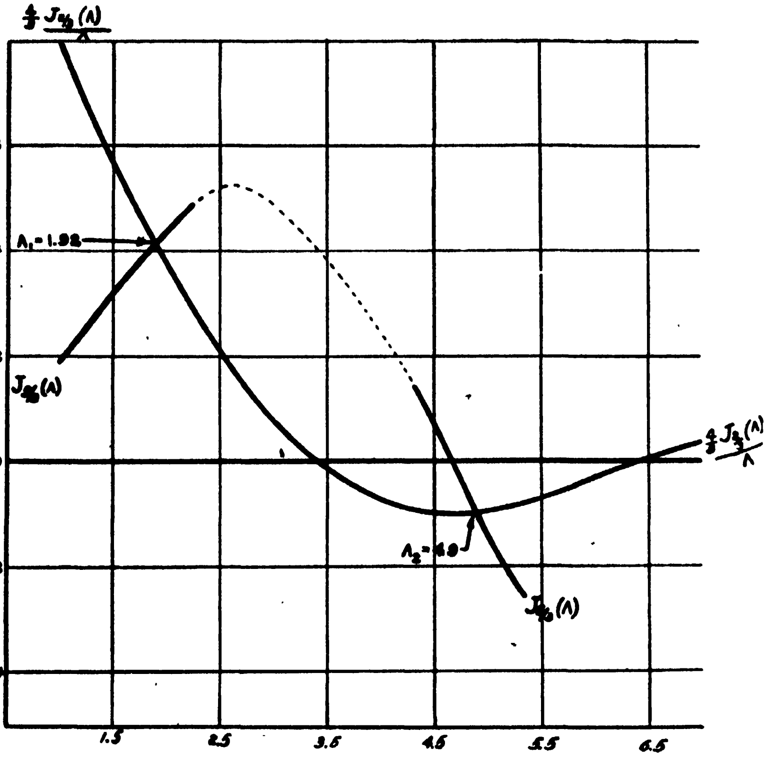


FIG. 2. Graphical solution for first two roots, $\Lambda_1$ and $\Lambda_2$ of Eq. (9).

$\partial E/\partial x$ must be zero for $z=1$, that is,

$$(\Lambda z^3)J_{2/3}'(\Lambda z^3) = -\tfrac{2}{3}J_{2/3}(\Lambda z^3) \quad \text{for } z=1.$$

Hence, the equivalent expression to be satisfied is:

$$J_{5/3}(\Lambda) = \frac{4}{3}\frac{J_{2/3}(\Lambda)}{(\Lambda)}. \tag{9}$$

The roots of this equation will determine the values of $\Lambda$ which will satisfy the boundary conditions at the base of the spring. Equation (9) may be solved graphically from tabulated values of $J_{5/3}(\Lambda)$ and of $4/3[J_{2/3}(\Lambda)/(\Lambda)]$. By inspection of Fig. 2, Eq. (9) is satisfied when $\Lambda = 1.92$, 4.9, $\cdots$,. Using the formulas developed in Section I, the natural frequencies may be computed. Table II shows the results of this case.

### IIB. Base Fixed, Vertex Free

In this case, the boundary conditions to be met are that $E(1, t) = 0$, and that $\partial E/\partial x$ be zero at $z=0$. It is advantageous at this point to write the general solution of Eq. (4) in the form:

$$W(y) = (y/2)^{\frac{1}{2}}\{AJ_{-2/3}(y) + BJ_{2/3}(y)\},$$

and noting that $y=0$ when $z=0$, $y=\Lambda$ when $z=1$. Now, from the definition of $J_{-2/3}(y)$, it is evident that

$$\lim_{y\to 0}[(y/2)^{\frac{1}{2}}J_{-2/3}(y)] = \frac{1}{\pi(-2/3)},$$

where $\pi(n) = \Gamma(n+1)$. Thus, when $y=0$, $E(z, t) = E(0, t) \neq 0$, as is required by boundary conditions, and the solution is not trivial. Inclusion of $J_{2/3}(y)$ in the solution is unnecessary to meet boundary conditions and $B$ can be taken as zero, thus reducing our solution to:

$$E(y, t) = (y/2)^{\frac{1}{2}}AJ_{-2/3}(y)\cos(\lambda t), \tag{10}$$

TABLE II.

| | Root | $\Lambda_i$ | $f_i$(c/sec.) | $f_i$(experimentally) |
|---|---|---|---|---|
| Ni | 1st | 1.92 | 1.53 | 1.53 |
| | 2nd | 4.9 | 3.9 | |
| | 3rd | 8.1 | 6.4 | |
| Piano wire | 1st | 1.92 | 1.49 | 1.51 |
| | 2nd | 4.9 | 3.8 | |
| | 3rd | 8.1 | 6.3 | |

or, in terms of $z$:

$$E(z, t) = [(\Lambda/2)^{\frac{1}{3}}A]z^2 J_{-2/3}(\Lambda z^3)\cos(\lambda t). \quad (11)$$

We now aim to satisfy the condition that $\partial E/\partial x$ be zero at $z=0$, using Eq. (11) as the particular solution of Eq. (1). This requires the following relationship to be satisfied:

$$3z(\Lambda z^3)J_{-2/3}'(\Lambda z^3) = -2zJ_{-2/3}(\Lambda z^3) \quad \text{for } z=0,$$

or the equivalent form:

$$3z[(-2/3)J_{-2/3}(\Lambda z^3) - \Lambda z^3 J_{1/3}(\Lambda z^3)] = -2zJ_{-2/3}(\Lambda z^3).$$

Terms in $J_{-2/3}(\Lambda z^3)$ are mutually annihilated and the following relationship remains to be satisfied:

$$-3\Lambda z^4 J_{1/3}(\Lambda z^3) = 0 \quad \text{for } z=0. \quad (12)$$

But this is true for $z=0$, hence, Eq. (11) satisfies the boundary condition at the vertex.

To satisfy the boundary condition at the base, $E(1, t) = 0$, the choice of $\Lambda$ must be restricted to roots of the equation $J_{-2/3}(\Lambda) = 0$. Hence, the boundary conditions of this case are satisfied by the particular solution (11) where $\Lambda_f$ are roots of $J_{-2/3}(\Lambda) = 0$. The first three roots of this equation are $\Lambda_1 = 1.25$, $\Lambda_2 = 4.43$, and $\Lambda_3 = 7.58$, while other roots may be deduced by adding $\pi$ successively to each former root.[2] After converting these roots to fundamental frequencies, the results of this case may be stated in tabular form (Table III).

Table III.

| | Root | $\Lambda_f$ | $f_f$(c/sec.) | $f_f$(experimentally) |
|---|---|---|---|---|
| Ni | 1st | 1.25 | 1.00 | 1.04 |
| | 2nd | 4.43 | 3.52 | |
| | 3rd | 7.58 | 6.04 | |
| Piano wire | 1st | 1.25 | .96 | .92 |
| | 2nd | 4.43 | 3.44 | |
| | 3rd | 7.58 | 5.88 | |

### IIIA. Weight Attached to Base, Vertex Fixed

The boundary conditions in this case are that $E(z, t)$ be zero at $z=0$ and that $K\partial E/\partial x = M\partial^2 E/\partial t^2$ at $z=1$, where $K = Gd^4/8z^3$, the spring "constant" in this problem, and $M$ the mass of the attached weight.

In order to satisfy the boundary condition at the vertex, $A$ in the general solution of Eq. (1) must be taken as zero. Hence,

$$(\Lambda/2)^{\frac{1}{3}}Bz^2 J_{2/3}(\Lambda z^3)\cos(\lambda t) \quad (13)$$

satisfies the physical conditions imposed at the point of support. To fulfill the boundary condition at $z=1$ (point of attachment of weight), the following equality must hold:

$$\frac{aGd^4}{4}\{4J_{2/3}(\Lambda) - 3\Lambda J_{5/3}(\Lambda)\} = 9Mc\Lambda^2 J_{2/3}(\Lambda),$$

or, writing this equation in a form more easily accessible to computation,

$$\left\{1 - \left[\frac{9Mc}{aGd^4}\right]\Lambda^2\right\} \cdot \frac{4}{3}\frac{J_{2/3}(\Lambda)}{(\Lambda)} = J_{5/3}(\Lambda). \quad (14)$$

Thus, the particular solution of Eq. (1) which satisfies the conditions of this case is Eq. (13) wherein $\Lambda_f$ are roots of Eq. (14). The defining equation for $\Lambda_f$, (14), and hence, for the fundamental frequencies of this system, may be solved by graphical means. The right member in Eq. (9) tabulated for Case IIA, is present here again. The function to the left of the equals sign in Eq. (14) is then built up as a tabulated function of $\Lambda$ and is plotted on the same sheet as $J_{5/3}(\Lambda)$. The intersection of these curves gives the roots $\Lambda_f$. Three different weights (60, 80, and 110 grams) are used in each case. For the coils used here, it was found that $aGd^4 = 5.22$ for the nickel coil, and for the piano wire coil, a corresponding value of $aGd^4 = 3.72$. Figure 3 gives, by inspection, the roots of Eq. (14) necessary for the fulfillment of the boundary conditions of this case. These roots are then converted into natural frequencies, yielding the following summary (Table IV).

### IIIB. Weight Attached to Vertex, Base Fixed

The boundary conditions in this case are that $E(z, t)$ be zero at $z=1$, and that $-K\partial E/\partial x = M\partial^2 E/\partial t^2$ at $z=0$, where $K = Gd^4/8z^3$, the spring "constant" in this problem, and $M$ the mass of the attached weights.

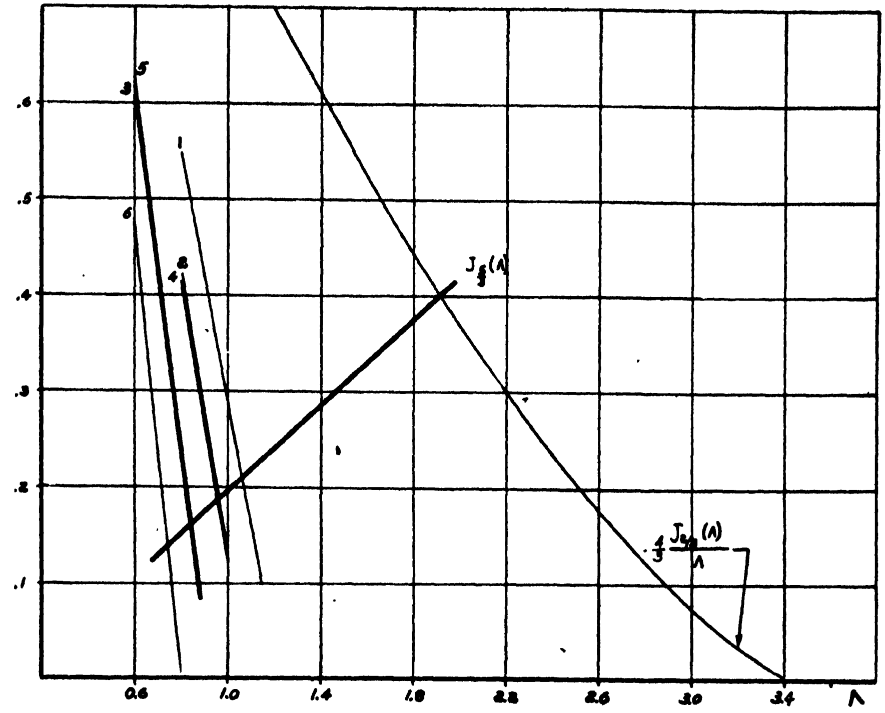


FIG. 3. Graphical solution of Eq. (14) for six values of $9Mc/aGd^4$.

To fulfill the boundary conditions of this problem, it is necessary to employ Bessel functions of orders $\frac{2}{3}$ *and* $-\frac{2}{3}$. The solution taken is the general one:

$$E(z,t)=(\Lambda/2)^{2/3}z^2$$
$$\times\{AJ_{-2/3}(\Lambda z^3)+BJ_{2/3}(\Lambda z^3)\}\cos(\lambda t). \quad (15)$$

The following two equalities must hold for the solution of this case:

$$\frac{aGd^4}{4z^2}(\Lambda/2)^{2/3}\{3A(\Lambda z^3)J_{-2/3}'(\Lambda z^3)+3B(\Lambda z^3)J_{2/3}'(\Lambda z^3)$$
$$+2AJ_{-2/3}(\Lambda z^3)+2BJ_{2/3}(\Lambda z^3)\}$$
$$=9Mc\Lambda^2(\Lambda/2)^{2/3}z^2\{AJ_{-2/3}(\Lambda z^3)+BJ_{2/3}(\Lambda z^3)\}, \quad (16)$$

for $z=0$, *and*

$$(\Lambda/2)^{2/3}z^2\{AJ_{-2/3}(\Lambda z^3)+BJ_{2/3}(\Lambda z^3)\}$$
$$\times\cos(\lambda t)=0 \quad (17)$$

for $z=1$. The condition of Eq. (17) reduces to:

$$AJ_{-2/3}(\Lambda)+BJ_{2/3}(\Lambda)=0. \quad (18)$$

In order to reduce Eq. (16) it is advisable to convert the derivatives of Bessel functions into related Bessel functions, and then to take limits of each term as $z\to 0$. Doing this, the condition to be met at the vertex becomes:

$$\left[\frac{aGd^4\Lambda^{2/3}}{(0.90275)4^{2/3}}\right]B_j=-\left[\frac{9Mc\Lambda^2 4^{2/3}}{(2.67893)\Lambda^{2/3}}\right]A_j. \quad (19)$$

Applying also condition (18), $A_j$ and $B_j$ can be eliminated, and the final ratio obtains:

$$\left\{-\left[\frac{aGd^4}{Mc}\right](59.4)\right\}\Lambda^{-2/3}=-\frac{J_{2/3}(\Lambda)}{J_{-2/3}(\Lambda)}, \quad (20)$$

where $M$ has been converted into grams. The values of $\Lambda_j$ which satisfy Eq. (20) are the ones which, when put into solution (15), will have both boundary conditions satisfied. Equation (20) may be solved graphically. $J_{2/3}(\Lambda)$ and

TABLE IV.

| Wire | Mass attached in grams | 1st frequency $\Lambda_1$ | $f_1$ | $f_1$(experimentally) |
|---|---|---|---|---|
| Nickel | 60 | 1.06 | .84 | .84 |
| | 80 | .96 | .76 | .74 |
| | 110 | .84 | .66 | .64 |
| Piano wire | 60 | .96 | .74 | .73 |
| | 80 | .85 | .66 | .64 |
| | 110 | .74 | .57 | .56 |

TABLE V.

| Wire | Mass attached in grams | 1st frequency $\Lambda_1$ | $f_1$ | $f_1$(experimentally) |
|---|---|---|---|---|
| Nickel | 60 | .89 | .71 | .71 |
|  | 80 | .82 | .66 | .65 |
|  | 110 | .74 | .59 | .58 |
| Piano wire | 60 | .81 | .62 | .61 |
|  | 80 | .75 | .58 | .56 |
|  | 110 | .66 | .51 | .50 |

$J_{-2/3}(\Lambda)$ have been tabulated against $\Lambda$ as independent variable. $aGd^4/Mc$ will have a constant value, depending on the spring used and the mass suspended. Figure 4 shows the final determination of the roots of Eq. (20). It gives, by inspection, the roots of Eq. (20) necessary for the fulfillment of the boundary conditions of this case. These roots may be converted into natural frequencies in the usual way. The following is a tabulation of the results of the theory and their experimental verification (Table V).

TABLE VI.

| Ratio of diameters | 1st root $\Lambda_1$ | nickel coil $f_1$ theory | nickel coil $f_1$ exp. | piano wire coil $f_1$ theory | piano wire coil $f_1$ exp. |
|---|---|---|---|---|---|
| Case I (1: ∞) | 3.38 | (see results of Case I) |  |  |  |
| $1:10^{\frac{1}{3}}$, (1:2.154) | 3.6 | 2.86 | 2.86 | 2.80 | 2.81 |
| $1:8^{\frac{1}{3}}$, (1:2.000) | 3.7 | 2.94 | 2.94 | 2.88 | 2.89 |
| $1:6^{\frac{1}{3}}$, (1:1.817) | 3.85 | 3.05 | 3.02 | 2.99 | 2.97 |

## IV. Two Intermediate Points Fixed

Next considered is the general case of the spring fastened at any two intermediate points. The

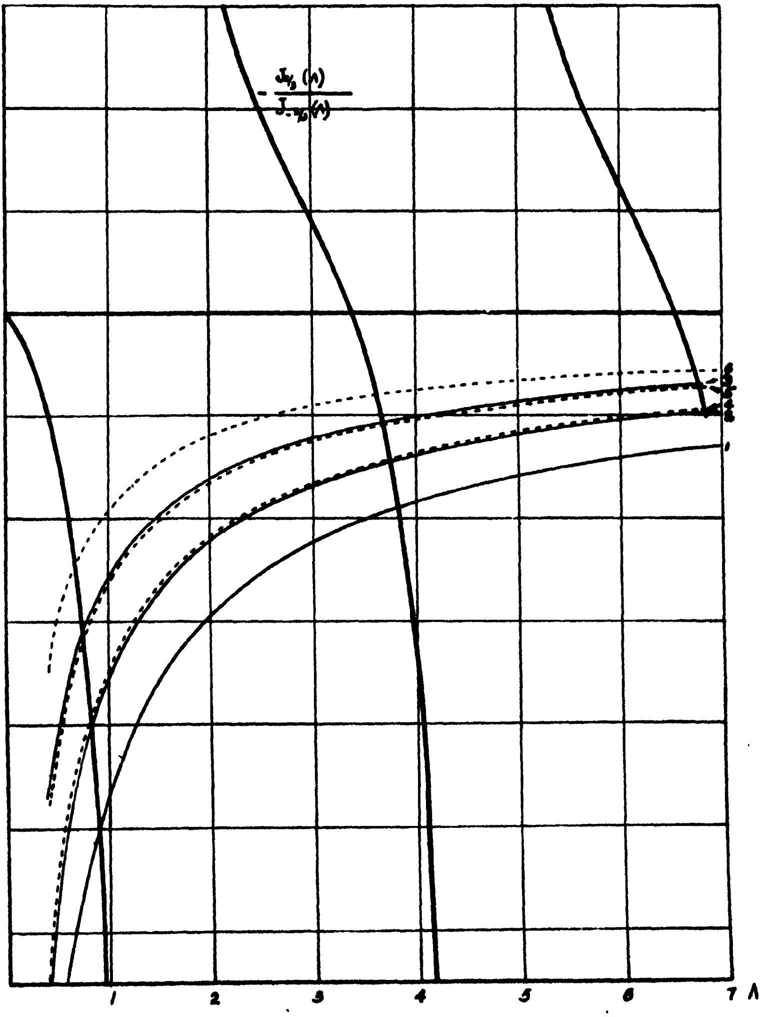


FIG. 4. Graphical solution of Eq. (20) for six values of $aGd^4/Mc$.

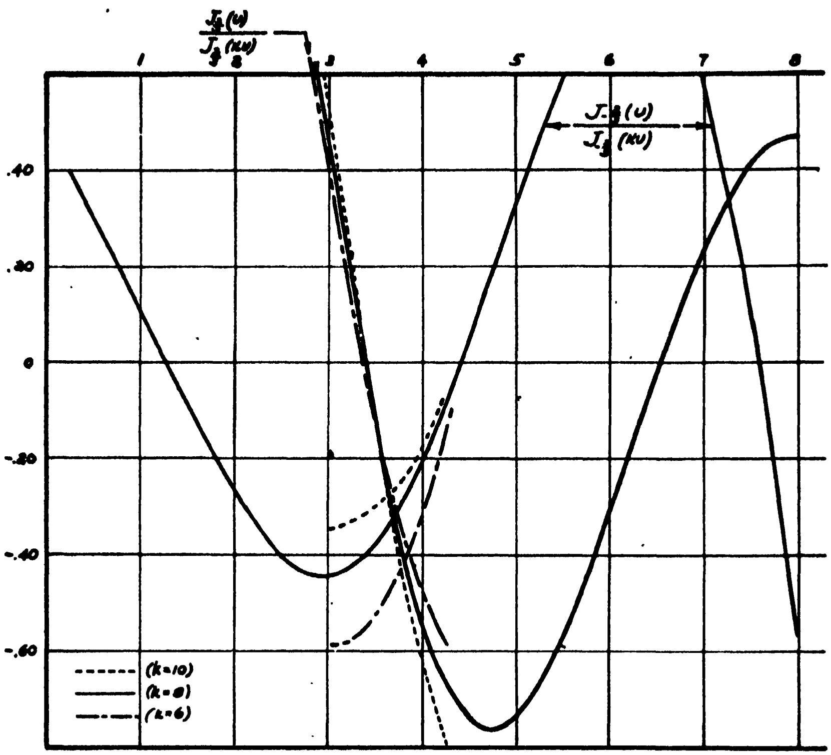


FIG. 5. Graphical solution of Eq. (24) when ratios of largest diameter of coil to smallest are as $1:10^{\frac{1}{3}}$, $1:2$, and $1:6^{\frac{1}{3}}$.

boundary conditions, stated mathematically, are that $E(z', t) = 0$ for $z' = a$ and for $z' = b$. There is no loss in generality if the widest upper boundary is taken to have the value 1. This can be done since in Eq. (1), $c$ is defined as $a^2Gd^4/2\pi mD_0^6$, where $D_0$ is the diameter of the widest turn of wire. This is, in effect, thinking of the upper bound as the base of the spring. Thus, the boundary conditions to be used are: $E(1, t) = 0 = E(b, t)$, where $0 < b < 1$. We take as a solution of Eq. (4) the general solution:

$$E(z,t) = (\Lambda/2)^{\frac{1}{3}} z^2 \{A J_{-2/3}(\Lambda z^3) + B J_{2/3}(\Lambda z^3)\} \times \cos(\lambda t). \qquad (21)$$

To satisfy boundary conditions at $z = 1$ and at $z = b$, the following equalities must be met simultaneously:

$$A J_{-2/3}(\Lambda) + B J_{2/3}(\Lambda) = 0, \qquad (22)$$

$$A J_{-2/3}(\Lambda b^3) + B J_{2/3}(\Lambda b^3) = 0. \qquad (23)$$

The solution of this pair of simultaneous equations depends upon the satisfaction of the following relationship:

$$\frac{J_{-2/3}(\Lambda)}{J_{-2/3}(\Lambda b^3)} = \frac{J_{2/3}(\Lambda)}{J_{2/3}(\Lambda b^3)}. \qquad (24)$$

Equation (24) may be solved graphically for specific values of $b^3$. Values of $J_n(k\Lambda_j)$ for $n = \frac{2}{3}$ and $-\frac{2}{3}$, and for $k = 6$, 8, and 10 are derived from tabulated values of $J_{-2/3}(x)$ and $J_{2/3}(x)$ by a shift of axis of the independent variable. It will be noted that this case degenerates into Case I when $k$ becomes infinite (that is, when the lower diameter is taken as zero). Figure 5 shows the solution of Eq. (24) for values of $b^3 = 6$, 8, and 10. It gives, by inspection, the values of $\Lambda_j$ which will satisfy Eq. (21) and both boundary conditions. $\Lambda_j$ may be translated into $f_j$ and in the usual way. Table VI below gives the results of this case.

## ACKNOWLEDGMENT

Special appreciation is due Professor D. A. Wells for having suggested this problem and for having offered many stimulating suggestions during its composition.

# An Introduction to the Mathematical Theory of Plasticity

W. PRAGER
*Brown University, Providence, Rhode Island*

(Received December 12, 1946)

The paper aims at introducing the reader to the principal theories of plasticity. Since a presentation of the general stress-strain relations used in these theories would require too much space, the discussion is restricted to the mechanical behavior of plastic materials under shear. Theories of plastic deformation (Hencky, Nadai) and theories of plastic flow (Saint Venant-Lévy-Mises, Prandtl-Reuss, Prager) are illustrated by the example of a prismatic bar under torsion.

## 1. INTRODUCTION

IN designing a structural member or a machine part the engineer, as a rule, uses formulas which are based on the theory of elasticity in a more or less rigorous manner. The highest stress predicted by these formulas is then compared with the maximum stress of the same type which the material can stand without failure. A so-called factor of safety is derived from this comparison. Now, in all but the very simplest cases the factor of safety obtained in this manner will differ considerably from the factor of safety defined as the ratio of the ultimate load to the design load. The reason for this is, of course, the fact that the stresses set up by the ultimate load will exceed the elastic limit of the material and that, consequently, the theory of elasticity is unable to predict these stresses. In order to obtain the second factor of safety, the only one with a physical meaning, the mechanical behavior of the material beyond the elastic limit must be taken into account. This behavior must also be taken into consideration in the study of manufacturing processes which aim at obtaining a favorable stress distribution under service conditions by setting up suitable initial stresses through previous plastic deformations. Another important field of application of the theory of plasticity is furnished by the phenomenon of creep. Under a constant load acting over a long period of time and at a high temperature a slow but steady increase of the deformation may take place. The problem of predicting the total deformation due to creep which may occur during the lifetime of a structure is of an obvious importance. In the cases mentioned so far, the plastic deformations must be rather small, as otherwise they would endanger the safety of the structure. This assumption of small deformations will no longer be justified in the study of technological processes like rolling or drawing. However, in an introduction to the theory of plasticity, it would certainly be unwise to try to cope at the same time with the difficulties arising from the consideration of finite deformations and from the unelastic behavior of the material. In the following it will, therefore, be assumed that the deformations, though greater than those possible within the elastic range, can still be treated as small. In order to show that such an assumption is reasonable, let us consider a bar of mild steel in simple tension. At the elastic limit the unit extension is of the order of one per mille. Now, we certainly are entitled to treat a unit extension of one percent still as small, but it is already far greater than any unit extension possible within the elastic range.

## 2. ELASTIC TORSION

The essential characteristic of the mechanical behavior of an elastic material is adequately expressed in Hooke's famous statement "*ut tensio sic vis.*" No similarly concise description of the mechanical behavior of plastic materials is possible, and a mathematical theory attempting to take into consideration all mechanical phenomena observed in the plastic range would not be practical. A workable mathematical theory must idealize the mechanical behavior of structural materials to a certain extent, and different idealizations are being used in the various fields of application. The fact that there exist almost as many theories of plasticity as writers on this subject is often a source of confusion to the student beginning to explore this field.

In the following, some of the more important theories of plasticity will be discussed briefly. Since a presentation of the general stress-strain relations used in the mathematical theory of plasticity would require too much space, the following discussion is restricted to the mechanical behavior of plastic materials under shear such as occurs at a generic point of a cylindrical or prismatic bar in torsion. Using a system of rectangular coordinates $x$, $y$, $z$, with the $z$ axis parallel to the generators of the cylindrical or prismatic surface of the bar, we let $\tau_x$ and $\tau_y$ denote the components of the shearing stress acting at a generic point of a cross section. We shall assume that all other stress components vanish and that $\tau_x$ and $\tau_y$ are independent of $z$.

The equation of equilibrium,[1]

$$\partial\tau_x/\partial x+\partial\tau_y/\partial y=0 \qquad (1)$$

is fulfilled if the stress components are derived from a stress function $\psi(x, y)$ in accordance with

$$\tau_x=\partial\psi/\partial y, \quad \tau_y=-\partial\psi/\partial x. \qquad (2)$$

The condition that the cylindrical or prismatic surface of the bar is free from stresses requires that the stress function $\psi$ has a constant value, for instance zero, along the boundary of the cross section of the bar.[1]

In order to obtain a differential equation for the stress function, we must consider the deformations of the bar and make use of the stress-strain relations. As in elastic torsion, the displacement components are assumed to be of the form

$$u=-yz\theta, \quad v=xz\theta, \quad w=\varphi(x, y, \theta), \qquad (3)$$

where $\theta$ denotes the angle of twist per unit length, and the function $\varphi$ represents the warping of an originally plane cross section of the bar.[2] The shear strains corresponding to the displacements (3) are[3]

$$\gamma_x=\partial\varphi/\partial x-y\theta, \quad \gamma_y=\partial\varphi/\partial y+x\theta. \qquad (4)$$

Up to this point, our discussion of the torsion problem applies equally well to elastic and plastic materials. Further progress, however, depends on the introduction of a stress-strain relation, and different materials require different stress-strain relations. For the purpose of the present discussion, the stress-strain relation of an *isotropic* material is best expressed as a relation between the vector $\boldsymbol{\tau}$ with the components $\tau_x$, $\tau_y$ and the vector $\boldsymbol{\gamma}$ with the components $\gamma_x$, $\gamma_y$, both vectors being considered as functions of time, as is the angle $\theta$. In addition to the vectors $\boldsymbol{\tau}$ and $\boldsymbol{\gamma}$, a stress-strain relation may therefore contain their derivatives with respect to time, $\partial\boldsymbol{\tau}/\partial t$, $\partial\boldsymbol{\gamma}/\partial t$, $\partial^2\boldsymbol{\tau}/\partial t^2$, etc. In expressing the stress-strain relation in vector form, we make certain that it is independent of the choice of the coordinate axes $x$, $y$, as it must be for an isotropic material.

For an elastic material, the stress-strain relation (Hooke's law) has the following particularly simple form

$$\boldsymbol{\tau}=G_0\boldsymbol{\gamma}, \qquad (5)$$

where $G_0$ is the shear modulus, a characteristic constant of the material. Introducing (2) and (4) into the scalar equations equivalent to (5), we obtain

$$\begin{aligned}\partial\psi/\partial y&=G_0(\partial\varphi/\partial x-y\theta),\\ -\partial\psi/\partial x&=G_0(\partial\varphi/\partial y+x\theta).\end{aligned} \qquad (6)$$

Elimination of the warping function $\varphi$ from Eqs. (6) furnishes the following differential equation for the stress function:

$$\partial^2\psi/\partial x^2+\partial^2\psi/\partial y^2=-2G_0\theta, \qquad (7)$$

where $G_0$ and $\theta$ may be considered as given constants. Together with the condition that $\psi=0$ on the boundary of the cross section, Eq. (7) completely defines the stress function $\psi$.

As L. Prandtl[4] has pointed out, the function $\psi$ satisfying (7) and vanishing along the boundary can be obtained by experiment. Let us consider a thin membrane, for instance a soap film, which is uniformly stretched over an opening in a rigid plane plate. The edge of the opening is given the shape of the contour of the cross section for which we desire to solve the boundary value problem formulated above. Furthermore, the membrane is fastened to the edge of the opening.

---

[1] See, for instance, S. Timoshenko, *Theory of Elasticity* (McGraw-Hill Book Company, Inc., New York, 1934) p. 230.

[2] J. Mandel, Ann. Ponts Chaussées **116**, 1–33 (1946) assumes $w=\theta\varphi(x, y)$. Such proportionality between warping and angle of twist can hardly be expected in the plastic range.

[3] See, for instance, S. Timoshenko, reference 1, p. 233.

[4] L. Prandtl, Jahresber. Deutsch. Math.-Ver. **13**, 31–36 (1904).

If we now have a small excess pressure $p$ on one side of the plate, the membrane will be deformed and it can easily be shown that the deflections $\omega$ of the membrane out of its original plane will satisfy the differential equation

$$\partial^2\omega/\partial x^2+\partial^2\omega/\partial y^2=-p/T, \qquad (8)$$

where $T$ is the uniform surface tension in the membrane. Along the edge of the opening we have $\omega=0$. Choosing the value of $p/T$ in a suitable manner, we can therefore obtain the surface $z=\psi(x, y)$ by experiment. Equations (2) show that the magnitude of the vector $\boldsymbol{\tau}$ is represented by the slope of the deformed membrane, and that the vector $\boldsymbol{\tau}$ is tangential to the contour lines of the deformed membrane.[5]

### 3. STRESS-STRAIN RELATIONS FOR PLASTIC MATERIALS UNDER SHEAR

In its initial phase, engineering research in plasticity was concerned with the limits of the elastic range rather than the phenomena occurring beyond this range. The condition that a given state of stress and strain is on the boundary of the elastic domain is expressed mathematically by the vanishing of a certain expression in the components of stress and strain (*yield condition*). For isotropic materials, this expression must be invariant with respect to rotation of the coordinate axes. Since Hooke's law holds up to the boundary of the elastic domain, all strain components which might appear in the yield condition can be expressed in terms of stress components. The yield conditions for an isotropic material then stipulates the vanishing of a certain stress invariant. In the torsion problem, where the state of stress at a point is defined by the vector $\boldsymbol{\tau}$, the only yield condition of this kind is

$$|\boldsymbol{\tau}|-k=0, \qquad (9)$$

where $|\boldsymbol{\tau}|$ denotes the *intensity of stress*, i.e., the magnitude of the vector $\boldsymbol{\tau}$, and $k$ the yield stress in pure shear.

The yield condition (9) establishes the limit up to which the stress distribution can be obtained from the theory of elasticity. To carry the stress analysis beyond this point, we must explore the stress-strain relations in the plastic range. It was already pointed out above that, for the torsion problem, a stress-strain relation reduces to a relation between the vectors $\boldsymbol{\tau}$ and $\boldsymbol{\gamma}$ and their derivatives with respect to time. We shall restrict the following discussion to stress-strain relations which do not involve time derivatives of an order higher than the first. Moreover, we shall neglect all viscosity effects. This means that we shall only admit stress-strain relations which are homogeneous in the time rates of stress and strain. In spite of these restrictions which are suggested by the results of experiments on metals such as steel, copper, and aluminum, we still retain considerable freedom in fitting stress-strain relations to experimental evidence.

With each *homogenous* state of stress and strain in a plastic body there is associated a certain *permanent strain*, i.e., the residual strain after the stress is reduced to zero by unloading the body. That part of the strain which disappears upon unloading is called the *elastic strain*. In the case of the torsion problem, we shall denote the elastic and permanent strains by $\boldsymbol{\gamma}'$ and $\boldsymbol{\gamma}''$, respectively. Experience shows that the elastic strain is related to the stress by Hooke's law:

$$\boldsymbol{\tau}=G_0\boldsymbol{\gamma}'. \qquad (10)$$

In establishing stress-strain relations for plastic materials, we must distinguish between *loading* and *unloading*. An infinitesimal change of the state of stress and strain in a plastic material is said to constitute loading if it involves a change of the permanent strain. On the other hand, during an infinitesimal change of stress and strain which constitutes unloading the permanent strain remains constant. In the case of the torsion problem, we have $\partial\boldsymbol{\gamma}/\partial t=\partial\boldsymbol{\gamma}'/\partial t$ for unloading, and Eq. (10) suggests

$$\partial\boldsymbol{\tau}/\partial t=G_0\partial\boldsymbol{\gamma}/\partial t \qquad (11)$$

as the stress-strain relation for unloading. Equation (11) is used in all theories of plasticity as the stress-strain relation for unloading; these theories therefore differ only in the stress-strain relations for loading. Moreover, even for loading the elastic strain is supposed to be related to the stress by Hooke's law, so that the task of establishing a stress-strain relation for a plastic

---

[5] For a comprehensive survey of the elastic torsion problem see T. J. Higgins, J. App. Phys. 10, 248–259 (1942).

material consists essentially in defining the manner in which the permanent strain, or its rate of change, depend on the stress. For the torsion problem, in particular, we must define the manner in which $\gamma''$ or $\partial\gamma''/\partial t$ depend on $\boldsymbol{\tau}$.

Let us first consider the case where $\gamma''$ is made to depend on $\boldsymbol{\tau}$ (*theories of plastic deformation*). The isotropy of the material requires then that these two vectors have the same direction. This may be expressed by the equation

$$G_0\gamma''=(\lambda-1)\boldsymbol{\tau}, \tag{12}$$

where $\lambda$ is a scalar factor of proportionality which, as a rule, will be a function of the coordinates $x$, $y$, and the time. To obtain a proper relation between $\gamma''$ and $\boldsymbol{\tau}$ we must supplement Eq. (12) by a relation between the intensities $|\gamma''|$ and $|\boldsymbol{\tau}|$:

$$|\boldsymbol{\tau}|=G_0f(|\gamma''|). \tag{13}$$

In the following, Eq. (13) will be called the *flow condition.* The flow condition enables us to eliminate the factor of proportionality, $\lambda$, from Eq. (12). In the following, we shall consider two kinds of flow conditions:

(a) *for perfectly plastic materials:* $G_0f=k=\text{const.}$,

(b) *for materials with strain-hardening:* for the small permanent strains to which this discussion is restricted, $f$ is a monotonically increasing function of its argument; it is then convenient to write the inverse of (13) in the form

$$G_0|\gamma''|=|\boldsymbol{\tau}|[g(|\boldsymbol{\tau}|/G_0)-1]. \tag{14}$$

The elastic strain is given by Eq. (10), the permanent strain by (12) and (13). Combining Eqs. (10) and (12), we obtain the following relation for the total strain $\gamma=\gamma'+\gamma''$:

$$G_0\gamma=\lambda\boldsymbol{\tau}. \tag{15}$$

In the case of a perfectly plastic material Eq. (15) together with the flow condition $|\boldsymbol{\tau}|=k$ constitute the stress-strain relations for loading. This stress-strain relation must, of course, be supplemented by Hooke's law, Eq. (5), which is supposed to be valid as long as $|\boldsymbol{\tau}|<k$. Stress-strain relations of this type were first discussed by H. Hencky.[6] In the case of a material with strain-hardening the form (14) of the flow condition enables us to eliminate $\lambda$ from Eq. (15). Since both $\gamma'$ and $\gamma''$ have the direction of $\boldsymbol{\tau}$, we have

$$G_0|\gamma''|=G_0[|\gamma|-|\gamma'|]=G_0|\gamma|-|\boldsymbol{\tau}|.$$

Equating the last member of this continued equation to the right-hand side of (14), we find

$$G_0|\gamma|=|\boldsymbol{\tau}|g(|\boldsymbol{\tau}|/G_0).$$

Comparing this to the equation obtained by taking absolute values on both sides of (15), we find

$$\lambda=g(|\boldsymbol{\tau}|/G_0).$$

The stress-strain relation for the loading of a material with strain-hardening therefore takes the form

$$G_0\gamma=\boldsymbol{\tau}g(|\boldsymbol{\tau}|/G_0). \tag{16}$$

Setting $G_0/g=G$, we may write (16) in the form $\boldsymbol{\tau}=G\gamma$, where $G$ is a function of $|\boldsymbol{\tau}|/G_0$. This stress-strain relation which, of course, applies to loading only, has the form of Hooke's law, but the shear modulus $G$, the so-called *secant modulus*, depends on the intensity of stress. The relation $\boldsymbol{\tau}=G\gamma$ covers the elastic as well as the plastic range, if $G$ is made to equal $G_0$ for $|\boldsymbol{\tau}|<k$. The practical stress analysis is simplified, however, if a single analytic function $G(|\boldsymbol{\tau}|/G_0)$ is used for both the elastic and the plastic range. To obtain good results one must, of course, choose a function which departs but little from the initial value $G_0$ as long as $|\boldsymbol{\tau}|<k$. On account of its mathematical simplicity, this type of stress-strain relation which was introduced by A. Nadai[7] is widely used in the modern literature;[8] it gives a continuous transition from the elastic to the plastic state.

The stress-strain relations of Hencky and Nadai are sometimes criticized because they let the stress depend on the instantaneous strain only and not on the history of straining as might

---

[6] H. Hencky, Zeits. f. angew. Math. Mech. **4**, 323–334 (1924).

[7] A. Nadai, *Plasticity, a Mechanics of the Plastic State of Matter* (McGraw-Hill Book Company, Inc., New York, 1931), p. 75.

[8] See, for instance, B. Finzi, Atti Ac. Sci. Torino, Cl. Sci. Fiz. Mat. Nat. **76**, 222–238 (1941); F. P. Cozzone, J. Aero. Sci. **10**, 137–151 (1943); A. A. Ilyushin, Prikladnaia Mat. Mekh. **7**, 245–272 (1943); L. M. Kachanov, *ibid.* **6**, 187–196 (1942); E. Melan, Oesterreich. Ing.-Arch. **1**, 14–21 (1946). W. R. Osgood, J. Aero. Sci. **11**, 213–226 (1944); W. W. Sokolovsky, J. Appl. Mech. **13**, A1–A10 (1946).

be expected for plastic materials. We shall return to this criticism after discussing some rival theories of plasticity.

Let us now consider the case where $\partial\boldsymbol{\gamma}''/\partial t$ is made to depend on $\boldsymbol{\tau}$ (*theories of plastic flow*). The isotropy of the material requires that these two vectors have the same direction, and we may write

$$G_0\partial\boldsymbol{\gamma}''/\partial t=\mu\boldsymbol{\tau}, \tag{17}$$

where $\mu$ is a factor or proportionality similar to the factor $\lambda$ appearing in (12). Combining Eqs. (10) and (17), we obtain the following relation for the total strain $\boldsymbol{\gamma}=\boldsymbol{\gamma}'+\boldsymbol{\gamma}''$:

$$G_0\partial\boldsymbol{\gamma}/\partial t=\partial\boldsymbol{\tau}/\partial t+\mu\boldsymbol{\tau}. \tag{18}$$

For a perfectly plastic material with the flow condition $|\boldsymbol{\tau}|=k$ or $\boldsymbol{\tau}\cdot\boldsymbol{\tau}=k^2$, we have $\boldsymbol{\tau}\cdot\partial\boldsymbol{\tau}/\partial t=0$. Scalar multiplication of each side of Eq. (18) by $\boldsymbol{\tau}$ yields therefore

$$\mu=G_0\boldsymbol{\tau}\cdot\partial\boldsymbol{\gamma}/k^2\partial t=G_0\dot{W}/k^2, \tag{19}$$

where $\dot{W}=\boldsymbol{\tau}\cdot\partial\boldsymbol{\gamma}/\partial t$ denotes the rate at which work is done. Substituting (19) into (18) we obtain

$$G_0\partial\boldsymbol{\gamma}/\partial t=\partial\boldsymbol{\tau}/\partial t+G_0\dot{W}\boldsymbol{\tau}/k^2 \text{ with } |\boldsymbol{\tau}|=k. \tag{20}$$

This stress-strain relation for loading must, of course, be supplemented by Hooke's law, Eq. (5), which is supposed to be valid as long as $|\boldsymbol{\tau}|<k$. Stress-strain relations of this type were first discussed by L. Prandtl[9] and E. Reuss.[10]

As has already been indicated above, the sharp distinction between the elastic and plastic ranges with separate stress-strain relations complicates any stress analysis considerably. It therefore seems desirable to modify Eq. (20) so as to obtain a gradual transition from the elastic to the plastic state and take account of strain-hardening at the same time. As the present author[11] has shown, this can be accomplished without difficulty. In fact, let us drop the condition $|\boldsymbol{\tau}|=k$ in (20). Scalar multiplication of each side of this equation by $\boldsymbol{\tau}$ then furnishes

$$G_0\dot{W}=\tfrac{1}{2}\dot{S}+G_0\dot{W}S/k^2, \tag{21}$$

where $S=\boldsymbol{\tau}\cdot\boldsymbol{\tau}=|\boldsymbol{\tau}|^2$ is a measure of the intensity of stress and the dot denotes differentiation with respect to time. Solving (21) for $\dot{W}$ and substituting into (20), we obtain

$$G_0\frac{\partial\boldsymbol{\gamma}}{\partial t}=\frac{\partial\boldsymbol{\tau}}{\partial t}+\frac{\dot{S}/k^2}{2(1-S/k^2)}\boldsymbol{\tau}. \tag{22}$$

Let us apply this stress-strain relation to the case of pure shear, where $\tau_y=\gamma_y=0$. Equation (22) then takes the form

$$G_0\dot{\gamma}_x=\dot{\tau}_x+\frac{\tau_x\dot{\tau}_x/k^2}{1-\tau_x{}^2/k^2}\tau_x=\frac{\dot{\tau}_x}{1-\tau_x{}^2/k^2}. \tag{23}$$

With the initial condition $\tau_x=0$ for $\gamma_x=0$, integration of (23) furnishes

$$\tau_x=k\tanh(G_0\gamma_x/k).$$

The graph of $\tau_x$ *vs.* $\gamma_x$ has the initial slope $G_0$ and the asymptote $\tau_x=k$. The stress-strain relation (23) thus provides a continuous transition from elastic behavior governed by Hooke's law to the flow of a perfectly plastic material.

Equation (23) is easily modified so as to include strain-hardening. To this end we write

$$G_0\frac{\partial\boldsymbol{\gamma}}{\partial t}=\frac{\partial\boldsymbol{\tau}}{\partial t}+\frac{1}{2}\left(\frac{G_0}{G^*}-1\right)\frac{\dot{S}}{S}\boldsymbol{\tau}, \tag{24}$$

where $G^*$ denotes the so-called *tangent modulus* which depends on the stress intensity $S$. Indeed, specializing as in (23), we obtain

$$G_0\dot{\gamma}_x=\dot{\tau}_x+\left(\frac{G_0}{G^*}-1\right)\frac{\tau_x\dot{\tau}_x}{\tau_x{}^2}\tau_x=\frac{G_0}{G^*}\dot{\tau}_x$$

or

$$d\tau_x/d\gamma_x=G.^*$$

From a graph of $\tau_x$ *vs.* $\gamma_x$ the tangent modulus $G^*$ can be found as a function of $\tau_x{}^2$; the quantity $G^*$ appearing in (24) is the same function of $S$. The stress-strain relation (24) was suggested in 1942 by H. J. Laning in an unpublished manuscript; more general stress-strain relations of this type are discussed by G. H. Handelman, C. C. Lin, and W. Prager.[12]

The first term on the right-hand side of Eq. (18) represents the rate of change of the elastic

---

[9] L. Prandtl, Proc. 1st Internat. Congr. Appl. Mech., Delft, 1924, pp. 43–54.

[10] E. Reuss, Zeits. f. angew. Math. Mech. 10, 266–274 (1930).

[11] W. Prager, Proc. 5th Internat. Congr. Appl. Mech., Cambridge, Massachusetts, 1938, pp. 234–237.

[12] G. H. Handelman, C. C. Lin, and W. Prager, Quart. App. Math. 4, 397–407 (1947).

strain. For perfectly plastic materials, it has sometimes been suggested to neglect the elastic strains at the side of the often much larger permanent strains. Equation (18) then takes the form

$$\partial\gamma/\partial t=\mu'\boldsymbol{\tau} \text{ with } |\boldsymbol{\tau}|=k. \tag{25}$$

Stress-strain relations of this type were the first to be used in the theory of plasticity, in the two-dimensional case by B. de Saint Venant,[13] and in the three-dimensional case by M. Lévy[14] and R. v. Mises.[15] As will be shown below, serious objections can be raised against the neglection of elastic strains.[16] In view of this fact, the amount of effort which has been spent on developing this particular theory of plasticity is surprising.[17]

## 4. PLASTIC TORSION

In applying the stress-strain relations of the preceding section to the torsion problem, let us first consider perfectly plastic materials. These possess a clearly defined yield point below which they obey Hooke's law (Hencky, Prandtl-Reuss) or are rigid (Saint Venant-Lévy-Mises); in the plastic range they flow under constant $|\boldsymbol{\tau}|$. Once the applied torque is sufficiently large to produce yielding, the cross section can be divided into elastic (or rigid) and plastic regions. In the latter we have $|\boldsymbol{\tau}|=k$ or, in terms of the stress function,

$$(\partial\psi/\partial x)^2+(\partial\psi/\partial y)^2=k^2. \tag{26}$$

As the applied torque increases the plastic regions expand until, finally, they cover the entire cross section.

Let us first consider this final stage of the torsion test. Here Eq. (26) is satisfied throughout the cross section. The left-hand side of this equation is the square of the slope of the tangent plane of the surface $z=\psi(x, y)$ with respect to the plane $z=0$. According to (26) the surface $z=\psi(x, y)$ therefore is a surface of constant slope. We can produce such a surface experimentally by heaping sand on a horizontal tray which has the shape of the cross section. Indeed, the slope of such a sand hill has a constant value corresponding to the angle of friction, and the boundary condition $\psi=0$ is enforced by the shape of the tray. For a bar of quadratic cross section, for instance, the surface $z=\psi(x, y)$ is a pyramid, for a rectangular cross section it has the form of a roof and for a circular cross section the form of a cone.

In the case of a perfectly plastic material, and for a torque which produces plastic flow throughout the entire test specimen, the stress distribution can thus be obtained without reference to any specific relation between stresses and strains. This fact is usually expressed by stating that, under these circumstances, the stress distribution in torsion is statically *determinate.*[18]

So far, we have assumed that the torque is sufficiently large to produce plastic flow throughout the entire test specimen. For smaller torques only part of the material will become plastic. According to the theories of Hencky and Prandtl-Reuss, the stress function $\psi$ will satisfy Eq. (7) [with $(\partial\psi/\partial x)^2+(\partial\psi/\partial y)^2<k^2$] in the elastic region of the cross section and Eq. (26) in the plastic region. Continuity of the stress vector $\boldsymbol{\tau}$ at the boundary between elastic and plastic regions requires continuity of the first derivatives of $\psi$. The complete problem for the stress function has therefore the following form: *to determine a function $\psi(x,y)$ with continuous first derivatives, which vanishes along the contour of the cross section and satisfies Eq. (7) with $(\partial\psi/\partial x)^2+(\partial\psi/\partial y)^2<k^2$ in part of the cross section and Eq. (26) in the rest.*[19] As Nadai[18] has pointed out, Prandtl's membrane analogy can easily be modified so as to enable us to solve this problem by experiment. Let us consider the square cross section as an example. A membrane is then stretched uniformly over a square opening in a rigid plate and fastened to the edge of this opening. Over the opening we erect a pyramidal surface of the slope which corresponds to the yield stress $k$. If now the membrane is deformed by slowly increasing pres-

---

[13] B. de Saint Venant, Comptes rendus **70**, 473–480 (1870).

[14] M. Lévy, Comptes rendus **70**, 1323–1325 (1870).

[15] R. v. Mises, Goettinger Nachrichten 582–592 (1913).

[16] See also W. Prager, J. App. Phys. **15**, 65–71 (1944).

[17] Excellent surveys of the Saint Venant-Mises theory of plasticity have been given by H. Geiringer, *Fondements mathématiques de la théorie des corps plastiques isotropes* (Gauthier-Villars, Paris, 1937), and S. G. Mikhlin in the chapter entitled "The mathematical theory of plasticity" of the book *Some New Methods in the Mechanics of Continuous Media* (Russian), (Izd. A. N. USSR, 1938), pp. 157–216.

[18] This was first pointed out by A. Nadai, Zeits. f. angew. Math. Mech. **3**, 442–454 (1923).

[19] For an ingenious manner of constructing particular solutions of this problem by an inverse method see V. V. Sokolovsky, Prikladnaia Mat. Mekh. **6**, 241–246 (1942).

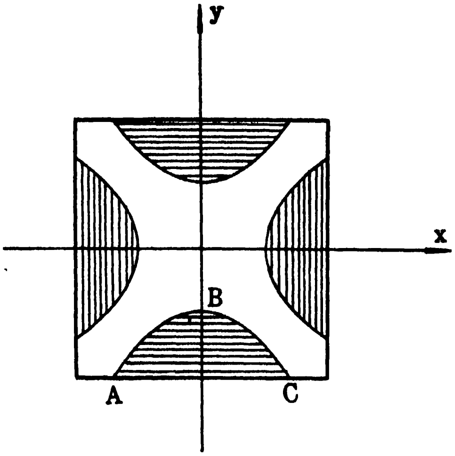


FIG. 1. Elastic and plastic regions in the cross section of a prismatic bar under torsion.

sure on one side of the plate, it will at first assume a shape which corresponds to the stress function in the elastic case. As the pressure increases, however, the membrane will partly be pressed against the pyramidal surface. Where this is the case, the membrane represents the plastic stress function satisfying Eq. (26). Where the membrane is not in contact with the pyramidal surface it represents the elastic stress function satisfying Eq. (7).

Having determined the stresses by means of this so-called soap film—sand hill analogy, we can determine the strains. Let us first apply the theory of Hencky, again considering a square cross section as an example. In Fig. 1, let the shaded areas be the plastic regions as obtained from the soap film—sand hill analogy and the unshaded cruciform area the elastic region. Since the stress function, and hence the stresses, are determined by the analogy, the strains in the elastic region can be found by applying Hooke's law. As Eqs. (4) show, this means that, in the elastic region, the warping function $\varphi$ can be determined to within an immaterial constant. To show how the warping function in the plastic region is found, we consider the shaded area $ABC$ in Fig. 1. Here $\tau_x = k$, $\tau_y = 0$, and the $y$ component of Hencky's stress-strain relation (15) gives

$$\gamma_y = \partial\varphi/\partial y + x\theta = 0, \qquad (27)$$

while the $x$ component does not furnish any information on account of the presence of the unknown factor $\lambda$. Integration of (27) furnishes

$$\varphi = -xy\theta + f(x). \qquad (28)$$

Along the arc $ABC$ the warping function $\varphi$ is known from the elastic solution, and this enables us to determine the function $f(x)$ in (28).

Next, let us apply the theory of Prandtl and Reuss to the same problem. The soap film—sand hill analogy shows that, for any given torque, this theory furnishes the same stress distribution as that of Hencky. Moreover, it can be shown that the warping function, too, is the same for both theories.[20] Indeed, let us consider a generic point $P$ of the cross section. As the torque is increased, the boundary between the elastic and plastic regions sweeps across the section of the bar. Once $P$ has been engulfed in the plastic region, the stress $\boldsymbol{\tau}$ at this point remains constant. For instance, if $P$ is engulfed in the shaded area $ABC$ of Fig. 1, the stress at $P$ will have the constant components $\tau_x = k$, $\tau_y = 0$. Accordingly, $\partial\boldsymbol{\tau}/\partial t$ is zero, and the $y$ component of the stress-strain relation (18) of the Prandtl-Reuss theory furnishes $\dot{\gamma}_x = 0$, while the $x$ component does not give any information on account of the presence of the unknown factor $\mu$. The strain component $\gamma_x$ therefore retains whatever value it has at the instant when the boundary between the elastic and plastic regions sweeps across $P$. On account of the continuity of the stress vector, however, the stress components at $P$ just before this point is engulfed into the plastic region must be $\tau_x = k$, $\tau_y = 0$. At this instant we therefore have $\gamma_y = 0$ at $P$, by Hooke's law. Since from this instant on $\dot{\gamma}_y = 0$ at $P$, we have $\gamma_y = 0$ at a generic point $P$ inside the shaded area $ABC$ of Fig. 1. From this and the identity of the stress distributions for the two theories follows the identity of the warping functions.

The theories of Hencky and Prandtl-Reuss are thus seen to furnish two identical predictions in the case of the torsion problem. This result must not be unduly generalized, however. As a rule, the predictions furnished by the two theories

---

[20] This was first pointed out by the present author in H. Geiringer and W. Prager, "Mechanik isotroper Koerper im plastischen Zustand," Ergebnisse d. exakt. Nat. wiss. 13, 310–363 (1934), p. 331.

differ; conditions under which they agree were investigated by A. A. Ilyushin.[21]

As regards the theory of Saint Venant-Lévy-Mises, it does not give a solution of the torsion problem except in the case where the torque is sufficiently large to produce plastic flow throughout the entire test specimen. Indeed, the stress distribution in the plastic part being the same as was studied above, Eq. (25) gives

$$\dot{\gamma}_y = \partial\dot{\varphi}/\partial y + x\dot{\theta} = 0$$

or

$$\dot{\varphi} = -xy\dot{\theta} + f(x) \tag{29}$$

for a point in the shaded region $ABC$ in Fig. 1. Since the cruciform unshaded region now is supposed to be rigid, we have $\dot{\varphi}=0$ along the arc $ABC$. This condition enables us to determine $f(x)$, and hence $\dot{\varphi}$, inside the shaded area $ABC$ provided that its boundary, the arc $ABC$, is known. *The theory of Saint Venant-Lévy-Mises does not furnish any condition, however, from which this boundary between the rigid and plastic regions could be determined. Moreover, the stress distribution in the rigid part is indeterminate.*

Let us review the results obtained so far. The Saint-Venant-Lévy-Mises theory is applicable only if the torque is large enough to produce plastic flow throughout the entire bar. For smaller torques the theories of Hencky and Prandtl-Reuss furnish identical results. In all but the very simplest cases, however, the theoretical determination of the boundary between the elastic and plastic regions encounters practically unsurmountable mathematical difficulties.[22] These difficulties are avoided by the theories of Nadai and Prager which stipulate a continuous transition from the elastic to the plastic state.

Let us first apply Nadai's theory to the torsion problem. In terms of the stress function the intensity of stress, on which the shear modulus depends, is given by

$$|\tau| = \left[\left(\frac{\partial\psi}{\partial x}\right)^2 + \left(\frac{\partial\psi}{\partial y}\right)^2\right]^{\frac{1}{2}} = |\text{grad}\psi|.$$

The stress-strain relation $\tau = G\gamma$ therefore gives

$$\frac{1}{G}\frac{\partial\psi}{\partial y} = \frac{\partial\varphi}{\partial x} - y\theta, \qquad -\frac{1}{G}\frac{\partial\psi}{\partial x} = \frac{\partial\varphi}{\partial y} + x\theta.$$

Elimination of $\varphi$ between these equations furnishes

$$\frac{\partial}{\partial x}\left(\frac{1}{G}\frac{\partial\psi}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{1}{G}\frac{\partial\psi}{\partial y}\right) = -2\theta \tag{30}$$

which equation replaces Eq. (7). Here $G$ is a function of $|\text{grad}\,\psi|$, and the stress function $\psi$ vanishes on the contour of the cross section. Equation (30) is non-linear, and somewhat reminiscent of equations which are encountered in connection with steady two-dimensional flows of compressible fluids. Integration methods developed in fluid dynamics might therefore prove useful; so far, however, the boundary value problem formulated in connection with Eq. (30) has been solved only in a few special cases.[23]

Finally, let us apply the stress-strain relation (24) to the torsion problem. To this end, we set

$$\int_0^S \frac{1}{2}\left(\frac{G_0}{G^*(s)} - 1\right)\frac{ds}{s} = -\log A(S).$$

This enables us to write Eq. (24) in the form

$$G_0\frac{\partial\gamma}{\partial t} = A\frac{\partial}{\partial t}(\tau/A). \tag{31}$$

Introducing the warping function and the stress function into the two scalar equations equivalent to (31) and eliminating the warping function between them we obtain[24]

$$\frac{\partial}{\partial x}\left[A\frac{\partial}{\partial t}\left(\frac{1}{A}\frac{\partial\psi}{\partial x}\right)\right] + \frac{\partial}{\partial y}\left[A\frac{\partial}{\partial t}\left(\frac{1}{A}\frac{\partial\psi}{\partial y}\right)\right] = -2G_0\frac{\partial\theta}{\partial t}. \tag{32}$$

This equation contains the time rates of change

---

[21] A. A. Ilyushin, Prikladnaia Mat. Mekh. 9, 207–218 (1945).

[22] For a numerical treatment of this problem by relaxation methods see D. G. Christopherson, Trans. A.S.M.E. 62, A1–A4, and F. S. Shaw, Austral. Counc. Aeronaut. Rep. ACA-11 (1944). See also R. V. Southwell, *Relaxation methods in theoretical physics* (Oxford Press, 1946).

[23] The as yet unpublished Ph.D. thesis of G. H. Handelman (Brown University, 1946) is concerned with the application of this theory to thin-walled sections. See also L. A. Galin, Prikladnaia Mat. Mekh. 8, 307–322 (1944).

[24] This equation was first established in W. Prager, "Theory of plasticity," (mimeographed lecture notes) Brown University, Providence, Rhode Island, 1942, p. 112.

of $\psi$ and $\theta$ in a homogeneous manner. Accordingly, the time scale is irrelevant and the angle of twist $\theta$ may be taken as measure for the time. Thus,

$$\frac{\partial}{\partial x}\left[A\frac{\partial}{\partial\theta}\left(\frac{1}{A}\frac{\partial\psi}{\partial x}\right)\right]+\frac{\partial}{\partial y}\left[A\frac{\partial}{\partial\theta}\left(\frac{1}{A}\frac{\partial\psi}{\partial y}\right)\right]=-2G_0. \quad (33)$$

Considered as an equation for $\psi(x, y, \theta)$, Eq. (33) is non-linear. It is worth noting, however, that this equation is linear in $\partial\psi/\partial\theta$ with coefficients which depend on $\psi$. This remark suggests the possibility of integrating (33) approximately by solving a number of linear problems for $\partial\psi/\partial\theta$ with coefficients which depend on the function $\psi$ built up from the previously obtained values of $\partial\psi/\partial\theta$. Physically speaking, this procedure amounts to producing the desired angle of twist in a series of small steps, and assuming that during each step the tangent modulus at any point of the cross section retains the value which corresponds to the stress intensity existing at the beginning of this step at the point under consideration.

For perfectly plastic materials, the theories of plastic deformation and of plastic flow furnished the same solution of the torsion problem; except for the circular cross section these theories will yield different results, however, when strain-hardening is taken into account. Torsion of a cylindrical bar of non-circular cross section may therefore afford an opportunity of deciding experimentally between these two theories.

---

# On the Theory of Noise in Radio Receivers with Square Law Detectors[*]

MARK KAC
*Department of Mathematics, Cornell University, Ithaca, New York*

AND

A. J. F. SIEGERT[**]
*Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Received September 13, 1946)

For the video output $V$ of a receiver, consisting of an i-f stage, a quadratic detector, and a video amplifier, the probability density $P(V)$ has been obtained for noise alone and for noise and signal. The results are expressed in terms of eigenvalues and eigenfunctions of the integral equation

$$\int_0^\infty K(t)\rho(s-t)f(t)dt=\lambda f(s),$$

where $\rho(\tau)$ is the i-f correlation function (i.e., the Fourier transform of the i-f power spectrum) and $K(t)$ is the response function of the video amplifier (i.e., the Fourier transform of the video amplitude spectrum). Two special cases are discussed in which the integral equation can be solved explicitly. Approximations for general amplifiers are given in the limiting cases of wide and narrow videos. Some applications of the method to other problems are shown in Sections 7B and 9.

## I. THE RECEIVER MODEL

THE response of a linear circuit to noise has been investigated by several authors. The probability densities and correlation coefficients of the noise outputs of some types of detectors are also known.[1] To find the probability density of the noise voltage output of the next stage

---

[*] The major part of this paper is based on work done for the Office of Scientific Research and Development under contract OEMsr-429 with Cornell University and contract OEMsr-262 with Massachusetts Institute of Technology.

[**] Now at the Physics Department, Syracuse University, Syracuse, New York.

[1] References to the work in this field can be found in the articles by S. O. Rice, Bell Sys. Tech. J. **23**, 282 (1944); **24**, 46 (1945), and M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1946).

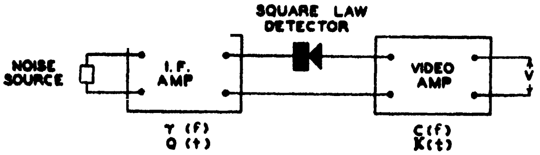


FIG. 1. The receiver model.

(video or audio amplifier) is a problem of a different type, since after the detector, the noise voltage is no longer Gaussianly distributed.

In this paper we have, therefore, calculated the probability density $P(V)$ of the noise voltage $V$ at the end terminals of a receiver consisting of an intermediate-frequency amplifier, a quadratic detector and a video (or audio) amplifier (Fig. 1). The response of both amplifiers is assumed to be linear and characterized by the steady-state response function $\gamma(f)$ and $C(f)$, respectively, or the response functions $Q(t)$ and $K(t)$, such that the output voltages of the amplifiers are $\gamma(f)e^{2\pi ift}$ and $C(f)e^{2\pi ift}$, respectively, if the input voltage is $e^{2\pi ift}$. If, therefore, the video input voltage is $v(t)$, the video output voltage $V$ is given by

$$V(t)=\int_{-\infty}^{\infty}dfC(f)e^{2\pi ift}\int_{-\infty}^{\infty}dt'e^{-2\pi ift'}v(t')$$
$$=\int_{-\infty}^{\infty}dt'K(t-t')v(t') \qquad (1.11)$$

where

$$K(\tau)=\int_{-\infty}^{\infty}dfC(f)e^{2\pi if\tau}. \qquad (1.12)$$

Since $V(t)$ can depend only on the values which $v(t')$ assumes at times $t'\leqq t$, it follows that $K(\tau)=0$ for $\tau<0$, which means that all poles of $C(f)$ lie above the real axis. Since $K$ must be real, we have $C(-f)=C^*(f)$. Both conditions are fulfilled for passive networks. We shall normalize $K$ by

$$\int_0^{\infty}K(t)dt=1.$$

In writing i-f voltages $V_{\text{I.F.}}^{(i)}$ and $V_{\text{I.F.}}^{(o)}$ for input and output it is generally convenient to consider them as modulated carriers, i.e., we write

$$V_{\text{I.F.}}^{(i)}=Re\{u(t)e^{2\pi if_0t}\}$$
$$=x(t)\cos 2\pi f_0t+y(t)\sin 2\pi f_0t,$$

where

$$u(t)=x(t)-iy(t),$$

and $f_0$ is the carrier frequency. The i-f output voltage is then

$$V_{\text{I.F.}}^{(o)}=Re\left\{\int_{-\infty}^{\infty}dfe^{2\pi ift}\gamma(f)\right.$$
$$\left.\times\int_{-\infty}^{\infty}dt'e^{-2\pi ift'}u(t')e^{2\pi if_0t'}\right\}$$
$$=Re\left\{e^{2\pi if_0t}\int_{-\infty}^{\infty}dt'u(t')\right.$$
$$\left.\times\int_{-\infty}^{\infty}dfe^{2\pi i(f-f_0)(t-t')}\gamma(f)\right\}$$
$$=Re\{U(t)e^{2\pi if_0t}\}$$

where

$$U(t)=\int_{-\infty}^{\infty}dt'Q(t-t')u(t'),$$

and

$$Q(\tau)=\int_{-\infty}^{\infty}dfe^{2\pi i(f-f_0)\tau}\gamma(f)=\int_{-\infty}^{\infty}df'e^{2\pi if'\tau}\gamma(f'+f_0).$$

We shall normalize $\gamma$ and $Q$ by demanding that

$$\int_{-\infty}^{\infty}Q^2(\tau)d\tau=1;$$

then we have

$$\int_{-\infty}^{\infty}Q^2(\tau)d\tau=\iiint_{-\infty}^{\infty}d\tau df'df''e^{2\pi i(f'+f'')\tau}\gamma(f'+f_0)$$
$$\times\gamma(f''+f_0)=1,$$

or

$$\int_{-\infty}^{\infty}df'\gamma(f_0+f')\gamma(f_0-f')=1.$$

The calculations are appreciably simplified by assuming a symmetrical pass band* for the i-f amplifier, i.e.,

$$\gamma(f_0+f')=\gamma^*(f_0-f'). \qquad (1.20)$$

We also put $|\gamma(f_0+f)|=B(f)$. Inserting this in the normalization equation we have

$$\int_{-\infty}^{\infty}|\gamma^2(f)|df=1,$$

i.e., the power spectrum is normalized. The sym-

---

* See, however, footnote 3, Section IV.

metry of $\gamma$ implies that $Q$ is real and $V_{\text{I.F.}}^{(o)}$ can now be written as

$$V_{\text{I.F.}}^{(o)}=X(t)\cos 2\pi f_0 t+Y(t)\sin 2\pi f_0 t \qquad (1.21)$$

with

$$\begin{aligned} X(t)&=\int_{-\infty}^{\infty}dt'Q(t-t')x(t') \\ Y(t)&=\int_{-\infty}^{\infty}dt'Q(t-t')y(t'). \end{aligned} \qquad (1.22)$$

$Q(\tau)$ must of course vanish for $\tau<0$.

The detector is assumed to yield the output

$$X^2(t)+Y^2(t), \qquad (1.31)$$

if the input voltage is

$$X(t)\cos 2\pi f_0 t+Y(t)\sin 2\pi f_0 t. \qquad (1.32)$$

In order that this be a reasonable model of an actual detector it is necessary to assume

$$\left|\frac{x'(t)}{x(t)}\right|\ll f_0 \quad\text{and}\quad \left|\frac{y'(t)}{y(t)}\right|\ll f_0,$$

which implies that the width of the pass band of the i-f amplifier is small compared to the intermediate frequency $f_0$.

## II. REPRESENTATION OF NOISE

The i-f output noise voltage (omitting the irrelevant d.c. component) is represented[2] as

$$V_{\text{I.F.}}^{(o)}=\frac{\sqrt{N}}{\sqrt{T}}\sum_{k=1}^{\infty}|\gamma(f_k')| \times(X_k'\cos 2\pi f_k't+Y_k'\sin 2\pi f_k't) \qquad (2.11)$$

with

$$f_k'=k/T,$$

where $X_1'$, $Y_1'$, $X_2'$, $Y_2'\cdots$ are independent normally distributed random variables each having the probability density

$$\pi^{-\frac{1}{2}}\exp(-u^2).$$

The representation holds only for the time interval $0<t<T$, but this is adequate if $T$ is chosen large compared to the time during which observations are made. We shall later let $T\rightarrow\infty$. The properties of $\gamma(f)$ were discussed above and $N$ denotes the input noise power per unit frequency.

According to (1.31) and (1.32), the detector output voltage arising from this i-f input voltage is given by

$$X^2(t)+Y^2(t)=V_D^{(o)},$$

where

$$\begin{aligned} X(t)&=\frac{\sqrt{N}}{\sqrt{T}}\sum_{k=1}^{\infty}|\gamma(f_k')|[X_k'\cos 2\pi(f_k'-f_0)t \\ &\qquad +Y_k'\sin 2\pi(f_k'-f_0)t] \\ Y(t)&=\frac{\sqrt{N}}{\sqrt{T}}\sum_{k=1}^{\infty}|\gamma(f_k')|[-X_k'\sin 2\pi(f_k'-f_0) \\ &\qquad +Y_k'\cos 2\pi(f_k'-f_0)t]. \end{aligned} \qquad (2.12)$$

Making use of (1.20) and the fact that the band width of the i-f amplifier is small compared to the intermediate frequency $f_0$, we may write

$$\begin{aligned} X(t)&=\frac{\sqrt{N}}{\sqrt{T}}\sum_{-\infty}^{\infty}B(f_k)(X_k\cos 2\pi f_k t+Y_k\sin 2\pi f_k t), \\ Y(t)&=\frac{\sqrt{N}}{\sqrt{T}}\sum_{-\infty}^{\infty}B(f_k) \\ &\qquad \times(-X_k\sin 2\pi f_k t+Y_k\cos 2\pi f_k t), \end{aligned} \qquad (2.13)$$

where

$$\begin{aligned} &f_k=k/T, \quad B(f_k)=|\gamma(f_k+f_0)|, \\ &X_k=X'_{k+k_0} \quad\text{and}\quad Y_k=Y'_{k+k_0} \end{aligned} \qquad (2.14)$$

with $k_0=Tf_0$, assumed to be an integer.

The video output voltage $V(t)$ is given by (1.11) as

$$\begin{aligned} V(t)&=\int_{-\infty}^{\infty}K(t-t')[X^2(t')+Y^2(t')]dt' \\ &=\int_0^{\infty}K(u)[X^2(t-u)+Y^2(t-u)]du. \end{aligned} \qquad (2.15)$$

In the derivation of Chapter 3 we shall consider the probability distribution of the slightly modified quantity

$$V_T(t)=\int_0^T K(u)[X^2(t-u)+Y^2(t-u)]du, \qquad (2.16)$$

and then pass to the limit as $T\rightarrow\infty$.

## III. "WHITE NOISE"

The i-f output $X(t)\cos 2\pi f_0 t+Y(t)\sin 2\pi f_0 t$ with $X(t)$, $Y(t)$ given by (2.13) can be described

---

[2] S. O. Rice, "Mathematical analysis of random noise," Bell Sys. Tech. J. 23, 282 (1944), see p. 328.

as caused by the "white noise"

$$x(t)\cos 2\pi f_0 t + y(t)\sin 2\pi f_0 t,$$

where $x(t)$ and $y(t)$ are given by

$$x(t)=\frac{\sqrt{N}}{\sqrt{T}}\sum_{k=-L}^{L}(X_k\cos 2\pi f_k t+Y_k\sin 2\pi f_k t)$$

$$v(t)=\frac{\sqrt{N}}{\sqrt{T}}\sum_{k=-L}^{L}(-X_k\sin 2\pi f_k t+Y_k\cos 2\pi f_k t),$$

where $L$ is a number so large that the frequencies $f_0 \pm L/T$ lie outside of the pass band of the i-f amplifier. This "white noise" is the input thermal and shot noise which actually is known to have a constant power spectrum up to very high frequencies.

In the derivation of the next section it will be advantageous to represent $x(t)$ and $y(t)$ in a different orthogonal system, and we will now show that the expansion coefficients in any other complete orthogonal system are Gaussianly distributed and independent of each other, if $L$ is sufficiently large, i.e., if the power spectrum is constant up to sufficiently high frequencies.

Let $\phi_j(t)$ be an orthonormal, complete set in the interval $(0, T)$. Expanding $x(t)$ in this system we have

$$x(t)=\sum_{1}^{\infty} u_j\phi_j(t), \tag{3.1}$$

where

$$u_j=\left(\frac{N}{T}\right)^{\frac{1}{2}}\sum_{k=-L}^{L}\left(X_k\int_0^T \phi_j(t)\cos 2\pi f_k t dt + Y_k\int_0^T \phi_j(t)\sin 2\pi f_k t dt\right). \tag{3.2}$$

The coefficients $u_j$ are thus linear combinations of the Gaussian variables $X_k$, $Y_k$ and are, therefore, Gaussianly distributed. To prove independence of linear combinations of independent Gaussian variables, it is necessary only to show that $\langle u_l u_j\rangle = 0$ for $l \neq j$. Using the statistical properties of $X_k$ and $Y_k$, we have

$$\langle u_l u_j\rangle_{\mathrm{Av}}=\frac{1}{2}\frac{N}{T}\int_0^T\int_0^T dsdt\phi_l(t)\phi_j(s) \times\sum_{-L}^{L}\cos 2\pi f_k(s-t). \tag{3.3}$$

For sufficiently large $L$

$$\frac{1}{T}\sum_{-L}^{L}\cos 2\pi f_k(s-t)$$

can be replaced by $\delta(s-t)$ and, therefore,

$$\langle u_l u_j\rangle_{\mathrm{Av}}\rightarrow\frac{N}{2}\delta_{jl}. \tag{3.4}$$

In the same way we can show, for the coefficients $v_j$ of $y(t)$, that

$$\langle v_l v_j\rangle_{\mathrm{Av}}\rightarrow\frac{N}{2}\delta_{lj} \tag{3.5}$$

and

$$\langle u_k v_l\rangle_{\mathrm{Av}}\rightarrow 0. \tag{3.6}$$

## IV. DERIVATION OF PROBABILITY DENSITY FOR NOISE ALONE

To derive the probability density $P(V)$ we use the above representation for the incoming "white noise" writing the i-f input voltage[3] as

$$V_{\mathrm{I.F.}}^{(i)}=x(t)\cos 2\pi f_0 t+y(t)\sin 2\pi f_0 t, \tag{4.11}$$

with

$$x(t)=\textstyle\sum_j u_j\phi_j(t)$$

and

$$y(t)=\textstyle\sum_j v_j\phi_j(t). \tag{4.12}$$

Choosing $N$ as unit of power we have

$$\langle u_k u_l\rangle_{\mathrm{Av}}=\langle v_k v_l\rangle_{\mathrm{Av}}=\tfrac{1}{2}\delta_{lk}. \tag{4.13}$$

The i-f output voltate $V_{\mathrm{I.F.}}^{(o)}$ is given by

$$V_{\mathrm{I.F.}}^{(o)}=X(t)\cos 2\pi f_0 t+Y(t)\sin 2\pi f_0 t, \tag{4.14}$$

where

$$X(t)=\int_{-\infty}^{\infty}Q(t-\tau)x(\tau)d\tau,$$
$$Y(t)=\int_{-\infty}^{\infty}Q(t-\tau)y(\tau)d\tau. \tag{4.15}$$

---

[3] The authors are aware that the results of this section can be obtained in a more elegant way by using the representation (4.12) for $V_{\mathrm{I.F.}}^{(i)}$ rather than for the components $x(t)$ and $y(t)$, and by defining the quadratic detector by

$$V_D^{(o)}=[V_{\mathrm{I.F.}}^{(o)}]^2$$

rather than by Eq. (4.31), provided that the I.F. and video band widths are small compared with $f_0$ [cf. Phys. Rev. 70, 449, C6 (1946)]. The simplifying assumption of symmetrical I.F. pass band can then be easily dispensed with. The method presented here was chosen to keep the continuity with the existing literature.

The correlation coefficient

$$\rho(\tau)=\langle X(t)X(t+\tau)\rangle_{\text{Av}}/\langle X^2(t)\rangle_{\text{Av}} \tag{4.21}$$

can be expressed in terms of the function $Q$ as follows:

$$\begin{aligned}\langle X(t)X(t+\tau)\rangle_{\text{Av}}&=\int\int_{-\infty}^{\infty} Q(t-\tau_1)Q(t+\tau-\tau_2)\\&\qquad\times\langle X(\tau_1)X(\tau_2)\rangle_{\text{Av}}d\tau_1d\tau_2\\&=\frac{1}{2}\int_{-\infty}^{\infty}Q(t-\tau_1)Q(t+\tau-\tau_1)d\tau_1\\&=\frac{1}{2}\int_{-\infty}^{\infty}Q(\theta)Q(\theta+\tau)d\theta,\end{aligned} \tag{4.22}$$

$$\langle X^2(t)\rangle_{\text{Av}}=\frac{1}{2}\int_{-\infty}^{\infty}Q^2(\theta)d\theta=\tfrac{1}{2}, \tag{4.23}$$

and therefore

$$\rho(\tau)=\int_{-\infty}^{\infty}Q(\theta)Q(\theta+\tau)d\theta. \tag{4.24}$$

We note that

$$\begin{aligned}&\int_{-\infty}^{\infty}e^{2\pi if\tau}\rho(\tau)d\tau\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}Q(\theta)Q(\theta+\tau)e^{2\pi if\tau}d\theta d\tau\\&=\int\int\int\int_{-\infty}^{\infty}df'df''\gamma(f'+f_0)\gamma(f''+f_0)\\&\qquad\times e^{2\pi i[f'\theta+f''(\theta+\tau)+f\tau]}d\theta d\tau\\&=\int\int_{-\infty}^{\infty}df'df''\gamma(f'+f_0)\gamma(f''+f_0)\\&\qquad\times\delta(f'+f'')\delta(f''+f)\\&=\gamma(f+f_0)\gamma(f_0-f)=|\gamma^2(f+f_0)|\end{aligned} \tag{4.25}$$

which is a special case of the Wiener-Khintchine relation.

The detector output voltage $V_D^{(o)}$ is, therefore,

$$\begin{aligned}V_D^{(o)}&=X^2(t)+Y^2(t)\\&=\int\int_{-\infty}^{\infty}Q(t-\tau_1)Q(t-\tau_2)\\&\qquad\times[x(\tau_1)x(\tau_2)+y(\tau_1)y(\tau_2)]d\tau_1d\tau_2,\end{aligned} \tag{4.31}$$

and the output of the video amplifier

$$\begin{aligned}V(t)&=\int_{-\infty}^{\infty}K(t-\theta)[X^2(\theta)+Y^2(\theta)]d\theta\\&=\int_{-\infty}^{\infty}K(t-\theta)d\theta\int\int_{-\infty}^{\infty}Q(\theta-\tau_1)Q(\theta-\tau_2)\\&\qquad\times[x(\tau_1)x(\tau_2)+y(\tau_1)y(\tau_2)]d\tau_1d\tau_2.\end{aligned} \tag{4.32}$$

Changing variables to

$$\theta'=t-\theta,\quad \tau_1'=t-\tau_1,\quad \tau_2'=t-\tau_2, \tag{4.33}$$

we have

$$\begin{aligned}V(t)=\int_{-\infty}^{\infty}&K(\theta')d\theta'\int\int_{-\infty}^{\infty}Q(\tau_1'-\theta')Q(\tau_2'-\theta')\\&\times[x(t-\tau_1')x(t-\gamma_2')\\&\qquad+y(t-\tau_1')y(t-\tau_2')]d\tau_1'd\tau_2'.\end{aligned} \tag{4.34}$$

If the i-f input noise voltage is a homogeneous random function $x(t-\tau)$ and $y(t-\tau)$ are given by

$$\begin{aligned}x(t-\tau)&=\textstyle\sum_j u_j'\phi_j(\tau),\\y(t-\tau)&=\textstyle\sum_j v_j'\phi_j(\tau),\end{aligned} \tag{4.35}$$

where $u_j'$ and $v_j'$ are random variables with the same properties as $u_j$ and $v_j$, and we have—omitting the primes—

$$\begin{aligned}V(t)&=\int_{-\infty}^{\infty}K(\theta)d\theta\int\int_{-\infty}^{\infty}Q(\tau_1-\theta)Q(\tau_2-\theta)\\&\qquad\times\textstyle\sum_{ij}(u_iu_j+v_iv_j)\phi_i(\tau_1)\phi_j(\tau_2)d\tau_1d\tau_2\end{aligned} \tag{4.36}$$

$$\begin{aligned}&=\textstyle\sum_{ij}(u_iu_j+v_iv_j)\\&\qquad\times\int\int_{-\infty}^{\infty}\Lambda(\tau_1,\tau_2)\phi_i(\tau_1)\phi_j(\tau_2)d\tau_1d\tau_2\end{aligned} \tag{4.37}$$

with

$$\Lambda(\tau_1,\tau_2)=\int_{-\infty}^{\infty}K(\theta)d\theta Q(\tau_1-\theta)Q(\tau_2-\theta). \tag{4.38}$$

We note that $t$ does not appear anymore on the r.h.s., that means that $V(t)$ is a homogeneous random function. This is caused by the fact that the input noise is itself a homogeneous random function.

The essential step in our calculation is now to choose as the function $\phi_i(\tau)$ the normalized eigenfunctions of the integral equation

$$\int_{-\infty}^{\infty} \Lambda(\tau_1, \tau_2)\phi(\tau_2)d\tau_2 = \lambda\phi(\tau_1). \qquad (4.41)$$

Then $V(t)$ is expressed as

$$V = \sum_i (u_i^2 + v_i^2)\lambda_i, \qquad (4.42)$$

where the quantities $\lambda_i$ are the eigenvalues of the integral equation. The quantities

$$u_i^2 + v_i^2 = Z_i \qquad (4.43)$$

are independent and their probability densities $W(Z_i)$ are given by

$$W(Z_i) = \begin{cases} e^{-Z_i} & \text{if } Z_i > 0, \\ 0 & \text{if } Z_i < 0, \end{cases} \qquad (4.44)$$

because $u_i$ and $v_i$ are independent Gaussian variables of spread $\frac{1}{2}$. The probability distribution of the sum representing $V$ is therefore

$$\begin{aligned} P(V)dV &= dV \int_0^{\infty} \cdots \int_0^{\infty} \delta(V - \sum_i \lambda_i Z_i) \prod_k W(Z_k) dZ_k \\ &= dV \frac{1}{2\pi} \int_{-\infty}^{\infty} d\xi e^{-iV\xi} \int_0^{\infty} \cdots \int_0^{\infty} \prod_k e^{-Z_k(1-i\lambda_k\xi)} dZ_k \\ &= dV \frac{1}{2\pi} \int_{-\infty}^{\infty} d\xi e^{-iV\xi} \frac{1}{\prod_k (1 - i\lambda_k \xi)}. \end{aligned} \qquad (4.45)$$

For $V>0$ the path of integration is closed in the lower half plane and contracted around all poles $\xi_s = -i/\lambda_s$ with $\lambda_s > 0$ and *vice versa*. We thus obtain

$$P(V) = \frac{1}{2\pi} \sum_s{}^{+} \oint^{(1/i\lambda_s, -)} \frac{d\xi}{(-i\lambda_s)\left(\xi - \frac{1}{i\lambda_s}\right)} \frac{e^{-iV\xi}}{\prod'_{k\neq s}(1 - i\lambda_k\xi)} = \sum_s{}^{+} \frac{e^{(-V/\lambda_s)}}{\lambda_s \prod'_{k\neq s}\left(1 - \frac{\lambda_k}{\lambda_s}\right)} \quad \text{for } V>0 \qquad (4.46)$$

where $\sum_s{}^{+}$ means summation over all values of $s$ for which $\lambda_s > 0$. For $V<0$ we obtain

$$P(V) = \frac{1}{2\pi} \sum_n{}^{(-)} \oint \frac{d\xi}{(-i\lambda_n)\left(\xi - \frac{1}{i\lambda_n}\right)} \frac{e^{-iV\xi}}{\prod'_{k\neq n}(1 - i\lambda_k\xi)} = \sum_n{}^{(-)} \frac{e^{(-V/\lambda_n)}}{\lambda_n \prod'_{k\neq n}\left(1 - \frac{\lambda_k}{\lambda_n}\right)} \qquad (4.47)$$

where $\sum_n{}^{(-)}$ indicates summation over all $n$ for which $\lambda_n < 0$. The product is extended always over all $k$ with exception of $k = n$.

It is sometimes more convenient to have the integral equation in a different form, involving the functions $K$ and $\rho$ rather than $K$ and $Q$. This form is obtained by introducing the new unknown

$$f(\theta) = \int_{-\infty}^{\infty} Q(\tau_2 - \theta)\phi(\tau_2)d\tau_2, \qquad (4.51)$$

so that

$$\int_{-\infty}^{\infty} d\theta K(\theta) Q(\tau_1 - \theta) f(\theta) = \lambda\phi(\tau_1). \qquad (4.52)$$

Multiplying both sides with $Q(\tau_1 - \theta')$ and integrating over $\tau_1$ we obtain

$$\int_{-\infty}^{\infty} K(\theta)d\theta \int_{-\infty}^{\infty} d\tau_1 Q(\tau_1 - \theta') \times Q(\tau_1 - \theta) f(\theta) = \lambda f(\theta'). \qquad (4.53)$$

We can now write

$$\int_{-\infty}^{\infty} Q(\tau_1 - \theta')Q(\tau_1 - \theta)d\tau_1 = \int_{-\infty}^{\infty} d\tau Q(\tau)Q(\tau + \theta' - \theta) = \rho(\theta' - \theta), \qquad (4.54)$$

and have the integral equation in the form

$$\int_{-\infty}^{\infty} K(\theta)\rho(\theta' - \theta) f(\theta) d\theta = \lambda f(\theta'),$$

or, since $K(\theta)=0$ for $\theta<0$,

$$\int_0^\infty K(\theta)\rho(\theta'-\theta)f(\theta)d\theta=\lambda f(\theta'). \quad (4.55)$$

Finally, by writing $(K(\theta))^{\frac{1}{2}}f(\theta)=\Phi(\theta)$ we can bring the equation into the form which will be derived directly in Section 6:

$$\int_0^\infty [K(\theta')]^{\frac{1}{2}}\rho(\theta'-\theta)[K(\theta)]^{\frac{1}{2}}\Phi(\theta)d\theta=\lambda\Phi(\theta'). \quad (4.56)$$

By means of the expression

$$V=\sum \lambda_j Z_j, \quad (4.61)$$

we can easily verify formulas for the moments. For the first two moments we have for instance

$$\langle V\rangle_{\text{Av}}=\sum \lambda_j\langle Z_j\rangle_{\text{Av}}=\sum \lambda_j=\int_0^\infty K(s)ds=1, \quad (4.62)$$

$$\begin{aligned}\langle V^2\rangle_{\text{Av}}&=\sum \lambda_j^2\langle Z_j^2\rangle_{\text{Av}}+\sum_{j\neq k}\lambda_j\lambda_k\langle Z_jZ_k\rangle_{\text{Av}}\\&=\sum_1^\infty \lambda_j^2+(\sum\lambda_j)^2 \qquad (4.63)\\&=1+\sum\lambda_j^2.\end{aligned}$$

For the fluctuation

$$\sigma^2=\langle V^2\rangle_{\text{Av}}-\langle V\rangle_{\text{Av}}^2. \quad (4.64)$$

We thus obtain

$$\sigma^2=\sum\lambda_j^2=\int_0^\infty\int_0^\infty K(s)K(t)\rho^2(s-t)dsdt, \quad (4.65)$$

this being the trace of the first iterated kernel.

## V. DERIVATION OF THE PROBABILITY DENSITY FOR SIGNAL AND NOISE

For the case of signal and noise only one derivation will be given here, since the direct derivation runs parallel to the analogous derivation for noise alone. The i-f input voltage $V_{\text{I.F.}}^{(i)}$ is in this case given by

$$V_{\text{I.F.}}^{(i)}=[x(t)+\alpha(t)]\cos 2\pi f_0t+[y(t)+\beta(t)]\sin 2\pi f_0t, \quad (5.11)$$

where $x(t)$, $y(t)$ are the input white noise amplitudes, and $\alpha(t)$, $\beta(t)$ are the input signal amplitudes. The i-f output voltage is then given by

$$V_{\text{I.F.}}^{(o)}=[X(t)+p(t)]\cos 2\pi f_0t+[Y(t)+q(t)]\sin 2\pi f_0t, \quad (5.12)$$

where $X$ and $Y$ are defined as above and

$$p(t)=\int_{-\infty}^\infty Q(t-\tau)\alpha(\tau)d\tau \quad\text{and}\quad q(t)=\int_{-\infty}^\infty Q(t-\tau)\beta(\tau)d\tau \quad (5.13)$$

are the signal amplitudes after the i-f amplifier. The detector output voltage is then

$$V_D^{(o)}=[X(t)+p(t)]^2+[Y(t)+q(t)]^2 \quad (5.14)$$

and the output of the video amplifier is

$$\begin{aligned}V(t)&=\int_{-\infty}^\infty K(t-\theta)d\theta\{[X(\theta)+p(\theta)]^2+[Y(\theta+q(\theta)]^2\}\\&=\int_{-\infty}^\infty K(t-\theta)d\theta\iint_{-\infty}^\infty Q(\theta-\tau_1)Q(\theta-\tau_2)\{[X(\tau_1)+\alpha(\tau_1)][x(\tau_2)+\alpha(\tau_2)]\\&\qquad+[y(\tau_1)+\beta(\tau_1)][y(\tau_2)+\beta(\tau_2)]\}d\tau_1d\tau_2. \quad (5.15)\end{aligned}$$

Changing variables to

$$\theta'=t-\theta,\ \tau_1'=t-\tau_1,\ \tau_2'=t-\tau_2, \quad (5.16)$$

we obtain

$$V(t)=\int_{-\infty}^{\infty}K(\theta')d\theta'\iint_{-\infty}^{\infty}Q(\tau_1'-\theta')Q(\tau_2'-\theta')\{[x(t-\tau_1')+\alpha(t-\tau_1')][x(t-\tau_2')+\alpha(t-\tau_2')]$$
$$+[y(t-\tau_1')+\beta(t-\tau_1')][y(t-\tau_2')+\beta(t-\tau_2')]\}d\tau_1'd\tau_2'$$
$$=\int_{-\infty}^{\infty}\Lambda(\tau_1,\tau_2)d\tau_1d\tau_2\{[x(t-\tau_1)+\alpha(t-\tau_1)][x(t-\tau_2)+\alpha(t-\tau_2)]$$
$$+[y(t-\tau_1)+\beta(t-\tau_1)][y(t-\tau_2)+\beta(t-\tau_2)]\}. \quad (5.17)$$

We now expand noise and signal amplitudes in terms of the eigen functions $\phi_j$ of $\Lambda$, writing

$$\begin{aligned} x(t-\tau)&=\textstyle\sum_j u_j\phi_j(\tau), & \alpha(t-\tau)&=\textstyle\sum_j \alpha_j(t)\phi_j(\tau),\\ y(t-\tau)&=\textstyle\sum_j v_j\phi_j(\tau), & \beta(t-\tau)&=\textstyle\sum_j \beta_j(t)\phi_j(\tau),\end{aligned} \quad (5.18)$$

and obtain

$$V(t)=\textstyle\sum_j\lambda_j\{[u_j+\alpha_j(t)]^2+[v_j+\beta_j(t)]^2\}. \quad (5.19)$$

The probability density of the variables $u_j$ and $v_j$ are $\pi^{-\frac{1}{2}}\exp(-u_j^2)$ and $\pi^{-\frac{1}{2}}\exp(-v_j^2)$, respectively, and the characteristic function for $V(t)$ is, therefore, given by

$$\langle e^{i\xi V}\rangle_{\text{Av}}=\int\overset{\infty}{\underset{-\infty}{\cdots}}\int\exp\{i\xi\textstyle\sum_j\lambda_j[(u_j+\alpha_j)^2+(v_j+\beta_j)^2]\}\displaystyle\prod_j\frac{\exp[-(u_j^2+v_j^2)]}{\pi}du_jdv_j$$

$$=\prod_j\frac{1}{\pi}\iint_{-\infty}^{\infty}du_jdv_j\exp\{-[(u_j^2+v_j^2)(1-i\xi\lambda_j)-2i\xi\lambda_j(u_j\alpha_j+v_j\beta_j)]+i\xi\lambda_j(\alpha_j^2+\beta_j^2)\}$$

$$=\prod_j\frac{1}{\pi}\iint_{-\infty}^{\infty}du_jdv_j\exp\left\{\left(-u_j(1-i\xi\lambda_j)^{\frac{1}{2}}-\frac{i\xi\lambda_j\alpha_j}{(1-i\xi\lambda_j)^{\frac{1}{2}}}\right)^2-\left(v_j(1-i\xi\lambda_j)^{\frac{1}{2}}-\frac{i\xi\lambda_j\beta_j}{(1-i\xi\lambda_j)^{\frac{1}{2}}}\right)^2\right\}$$
$$\times\exp\left[-\frac{\lambda_j^2\xi^2(\alpha_j^2+\beta_j^2)}{1-i\xi\lambda_j}+i\xi\lambda_j(\alpha_j^2+\beta_j^2)\right], \quad (5.21)$$

$$\langle e^{i\xi V}\rangle_{\text{Av}}=\prod_j\frac{\exp[i\xi\lambda_j(\alpha_j^2+\beta_j^2)/(1-i\xi\lambda_j)]}{1-i\xi\lambda_j}. \quad (5.22)$$

The Fourier inversion integral of this expression which represents the probability density cannot be evaluated in closed form.

We shall check this expression by deriving the known [4] probability density for the voltage after a quadratic detector. This distribution is obtained in the limit of infinitely wide video band width, which implies that

$$\int_0^{\infty}K(\theta)\psi(\theta)d\theta=\psi(0) \quad (5.31)$$

for any function $\psi(\theta)$. Then $\Lambda(\tau_1\tau_2)$ becomes simply the product $Q(\tau_1)Q(\tau_2)$ and the integral equation reduces to

$$Q(\tau_1)\int_{-\infty}^{\infty}Q(\tau_2)\phi(\tau_2)d\tau_2=\lambda\phi(\tau_1). \quad (5.32)$$

The only eigenfunction of this integral equation is of the form $CQ(\tau_1)$ where $C$ is a constant, and from

---

[4] S. A. Goudsmit, RL 43-21; K. A. Norton and V. D. Landon, Proc. I.R.E. **30**, 425, Sept. 42; J. C. Slater, RL V-23; D. G. Fink, Report T8; D. O. North, RCA Tech. Rep. PTR6C.

$\int_{-\infty}^{\infty}Q^2(\tau)d\tau=1$ we have $C=1$ and $\lambda=1$. The characteristic function, therefore, reduces to

$$\frac{\exp i\xi(\alpha^2+\beta^2)/(1-i\xi)}{1-i\xi},$$

where

$$\begin{aligned}\alpha&=\int_{-\infty}^{\infty}\alpha(t-\tau)Q(\tau)d\tau=\int_{-\infty}^{\infty}\alpha(\tau')Q(\tau-\tau')d\tau'=p(t),\\ \beta&=\int_{-\infty}^{\infty}\beta(t-\tau)Q(\tau)d\tau=\int_{-\infty}^{\infty}\beta(\tau')Q(t-\tau')d\tau'=q(t).\end{aligned}\tag{5.33}$$

Calling the signal after the detector $s^2=s^2(t)$ we have

$$s^2(t)=p^2(t)+q^2(t)=\alpha^2+\beta^2,\tag{5.34}$$

and

$$\begin{aligned}P(V)&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\exp(-i\xi V+i\xi s^2/(i-i\xi))}{(1-i\xi)}d\xi\\ &=\frac{1}{2\pi}\oint^{-i,-}\frac{\exp(-i\xi V+i\xi s^2/(1-i\xi))}{1-i\xi}d\xi\end{aligned}\tag{5.35}$$

for $V>0$ and zero otherwise. With $u=1-i\xi$ this becomes

$$\begin{aligned}P(V)&=\exp(-V-s^2)\frac{1}{2\pi i}\int^{0,+}\exp[Vu+s^2/u]du/u\\ &=\exp(-V-s^2)J_0(2is\sqrt{V})\text{ for }V>0\text{ and}\end{aligned}\tag{5.36}$$

zero otherwise, where $J_0$ is the Bessel function of order 0. This result is in agreement with the earlier results.

## VI. DIRECT DERIVATION FOR NOISE ALONE

In order to give a direct derivation of the results obtained in Section 4, we investigate the probability density of the function

$$V(t)=\int_0^T K(u)[x^2(t-u)+y^2(t-u)]du\tag{6.11}$$*

and let, at an appropriate instant, $T$ approach infinity. For the sake of simplicity, we shall restrict ourselves to the case $K(u)>0$. We divide the interval $(0, T)$ into a large number $n$ of equal subintervals and consider the expression

$$\begin{aligned}&\frac{T}{n}\sum_{j=1}^{n}[x^2(t-u_j)+y^2(t-u_j)]K(u_j);\\ &\frac{j-1}{n}T<u_j<\frac{j}{n}T.\end{aligned}\tag{6.12}$$

The characteristic function of the joint probability density of

$$x_1=x(t-u_1),\quad x_2=x(t-u_2),\quad\cdots,\quad x_n=x(t-u_n)$$

defined as the average

$$\left\langle\exp\left[i\sum_1^n\xi_jx(t-u_j)\right]\right\rangle_{\mathrm{Av}}$$

is easily found to be

$$\exp\left[-\frac{1}{4}\sum_{j,k=1}^{n}\rho_T(u_j-u_k)\xi_j\xi_k\right]\tag{6.13}$$

where the correlation coefficients $\rho_T$ for the finite interval $T$ are defined as

$$\rho_T(t)=\frac{1}{T}\sum_{-\infty}^{\infty}A(f_k)\cos 2\pi f_kt,\tag{6.14}$$

and

$$A(f)=B^2(f)=|\gamma^2(f\pm f_0)|.\tag{6.15}$$

---

* For convenience, $x$ and $y$ are used in this section to denote the quantities formerly denoted by $X$ and $Y$.

Thus the joint probability density of

$$x_1,\ x_2,\ x_3,\ \cdots,\ x_n$$

is given by the formula

$$\frac{1}{(2\pi)^n}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left\{-i\sum_1^n \xi_j x_j\right\}$$

$$\times\exp\left\{-\frac{1}{4}\sum_1^n \rho_T(u_j-u_k)\xi_j\xi_k\right\}d\xi_1\cdots d\xi_n$$

$$=\frac{1}{(\sqrt{\pi})^n}\frac{1}{\sqrt{D}}\exp\left\{-\sum_1^n \alpha_{jk}x_jx_k\right\},\quad (6.16)$$

where the matrix $[\alpha_{jk}]$ is the inverse of the correlation matrix $[\rho_T(u_j-u_k)]$ and $D$ the determinant of the correlation matrix.

We next find the characteristic function of

$$\frac{T}{n}\sum_1^n K(u_j)x^2(t-u_j),$$

i.e., the average

$$\left\langle\exp\left\{\frac{i\xi T}{n}\sum_1^n K(u_j)x^2(t-u_j)\right\}\right\rangle_{\text{Av}}.$$

We have

$$\left\langle\exp\left\{\frac{i\xi T}{n}\sum_1^n K(u_j)x^2(t-u_j)\right\}\right\rangle_{\text{Av}}$$

$$=\frac{1}{(\sqrt{\pi})^n}\frac{1}{\sqrt{D}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left\{\frac{i\xi T}{n}\sum_1^n K(u_j)x_j^2\right\}$$

$$\times\exp\left\{-\sum_1^n \alpha_{jk}x_jx_k\right\}dx_1\cdots dx_n.\quad (6.21)$$

Since $K(u_j)>0$ we can introduce the new variables

$$x_j'=[K(u_j)]^{\frac{1}{2}}x_j.\quad (6.22)$$

Obtaining

$$\left\langle\exp\left\{\frac{i\xi T}{n}\sum_1^n K(u_j)x^2(t-u_j)\right\}\right\rangle_{\text{Av}}$$

$$=\frac{1}{(\sqrt{\pi})^n}\frac{1}{\sqrt{\ }}\frac{1}{\prod_1^n[K(u_j)]^{\frac{1}{2}}}$$

$$\times\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left\{\frac{i\xi T}{n}\sum_1^n x_j'^2\right\}$$

$$\times\exp\left\{-\sum_1^n \beta_{jk}x_j'x_k'\right\}dx_1'\cdots dx_n',\quad (6.23)$$

where

$$\beta_{jk}=\frac{\alpha_{jk}}{[K(u_j)]^{\frac{1}{2}}[K(u_k)]^{\frac{1}{2}}}.\quad (6.24)$$

Let $\mu_1^{(n)},\ \mu_2^{(n)},\ \cdots,\ \mu_n^{(n)}$ be the eigenvalues of the matrix $[\beta_{jk}]$; then changing the coordinates in the last integral to "principal axes" of the "ellipsoid"

$$\sum_1^n \beta_{jk}x_j'x_k'=1\quad (6.25)$$

and noting that $\sum x_j'^2$ remains invariant under the transformation, we get easily

$$\left\langle\exp\left\{\frac{i\xi T}{n}\sum_1^n K(u_j)x^2(t-u_j)\right\}\right\rangle_{\text{Av}}$$

$$\frac{1}{\sqrt{D}\prod_1^n[K(u_j)]^{\frac{1}{2}}}\frac{1}{\prod_1^n\left(\mu_j^{(n)}-\frac{i\xi T}{n}\right)^{\frac{1}{2}}}\quad (6.26)$$

Since we have assumed that $A(f)$ is an even function we can show without much difficulty that

$$\sum_1^n K(u_j)x^2(t-u_j)\quad\text{and}\quad\sum_1^n K(u_j)y^2(t-u_j)$$

are independent in the statistical sense of the word. Thus the characteristic function of

$$\frac{T}{n}\sum_1^n K(u_j)\{x^2(t-u_j)+y^2(t-u_j)\}\quad (6.27)$$

is

$$\frac{1}{D\prod_1^n K(u_j)\prod_1^n\left(\mu_j^{(n)}-\frac{i\xi T}{n}\right)}$$

we now observe that the matrix $[\beta_{jk}]$ is the inverse of the matrix

$$[(K(u_j))^{\frac{1}{2}}\rho_T(u_j-u_k)(K(u_k))^{\frac{1}{2}}]\quad (6.28)$$

and that the determinant of (6.28) is

$$D\prod_1^n K(u_j).$$

Denoting by $\lambda_1^{(n)},\ \lambda_2^{(n)},\ \cdots,\ \lambda_n^{(n)}$, the eigenvalues of matrix (6.28) we see that the $\mu$'s are the inverses of the $\lambda$'s and that

$$D\prod_1^n K(u_j)=\lambda_1^{(n)}\cdot\lambda_2^{(n)}\cdots\lambda_n^{(n)}.$$

These facts imply that the characteristic function of (6.27) is

$$1\Big/\prod_{1}^{n}\left(1-\frac{iT}{n}\lambda_j^{(n)}\xi\right). \tag{6.29}$$

It follows from Hilbert's treatment of integral equations that the eigenvalues of matrix (6.28) when multiplied by $T/n$ approach, as $n$ approaches infinity, the eigenvalues of the integral equation

$$\int_0^T [K(s)]^{\frac{1}{2}}\rho_T(s-t)[K(t)]^{\frac{1}{2}}f(t)dt=\lambda f(s), \tag{6.31}$$

provided these eigenvalues are simple. Since the distribution of (6.27) approaches that of $V(t)$ we deduce that the characteristic function of $V(t)$ is given by the formula

$$1\Big/\prod_{1}^{\infty}(1-i\lambda_j^{(T)}\xi), \tag{6.32}$$

where $\lambda_1^{(T)}$, $\lambda_2^{(T)}$, $\cdots$, are the eigenvalues of the integral equation (6.31). For large $T$ we can replace the integral equation (6.31) by

$$\int_0^\infty [K(s)]^{\frac{1}{2}}\rho(s-t)[K(t)]^{\frac{1}{2}}f(t)dt=\lambda f(s), \tag{6.33}$$

and the characteristic function of $V(t)$ by

$$1/\prod_j(1-i\lambda_j\xi), \tag{6.34}$$

where $\lambda_1$, $\lambda_2$, $\lambda_3$, $\cdots$ are now the eigenvalues of the integral equation (6.34) which we again assume to be simple. The probability density $P(V)$ of $V$ can now be expressed by means of the inversion formula

$$P(V)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-iV\xi}}{\prod_j(1-i\lambda_j\xi)}d\xi. \tag{6.35}$$

At the end of Section 4, the integration has been carried out by the method of residues. The justification of the formal integration is not quite simple since one must investigate the behavior of the entire function

$$\phi(Z)=\prod_{1}^{\infty}(1-i\lambda_j Z)$$

for large $|Z|$.[5] However, it can be considerably simplified if one assumes that for some $\nu(0<\nu<1)$ the series

$$\sum_{1}^{\infty}\lambda_j^{\nu}$$

converges. On the other hand it is easy to see that the exact probability density $P(V)$ is the limit, as $n$ approaches infinity, of $P_n(V)$ where

$$P_n(V)=\sum_{s=1}^{n}\frac{\exp(-V/\lambda_s)}{\lambda_s\prod_{1}^{n}{}'(1-\lambda_j/\lambda_s)};\quad V>0. \tag{6.36}$$

$P_n(V)=0$; $V<0$ since with the assumption $K(u)>0$ negative eigenvalues cannot occur. In fact, the characteristic function corresponding to $P_n(V)$ which is seen to be

$$\prod_{1}^{n}\frac{1}{1-i\xi\lambda_j}$$

approaches the characteristic function associated with $P(V)$ thus $P_n(V)$ approaches $P(V)$.

## VII. EXPLICIT SOLUTIONS IN SOME SPECIAL CASES

### 1. Single-Tuned I.F. with Simple Low Pass Video

For the special case of a single tuned i-f amplifier with the video amplifier acting as a simple low pass filter the integral equation can be solved explicitly and the eigenvalues can be expressed in terms of roots of Bessel functions. For one special ratio of video band width to i-f band width it is even possible to obtain $P(V)$ in closed form. For other ratios it is not too difficult to sum the series for $P(V)$ numerically.

The correlation function $\rho(t)$ is in this case

$$\rho(t)=e^{-\alpha|t|} \tag{7.11}$$

which is equivalent to writing

$$A(f)=\frac{\text{Const.}}{1+(2\pi f/\alpha)^2}. \tag{7.12}$$

---

[5] We note that $\phi(Z)=D(iZ)$ where $D(\mu)$ is the Fredholm Determinant of the above integral equation. It is noteworthy that this relationship between the characteristic function of $P(V)$ and the Fredholm Determinant persists, even in the general case in which $K(t)$ is not assumed positive (in which case the simple product representation need not hold, since $\Sigma\lambda$ need not converge). In the general case one must take the Fredholm Determinant of the kernel $K(t)\rho(s-t)$.

For the video we assume

$$K(u)=\begin{cases}0 & , u>0\\ \beta\exp(-\beta u), & u<0\end{cases} \qquad (7.13)$$

i.e.,

$$C(f)=\frac{1}{1+2\pi i f/\beta}. \qquad (7.14)$$

The integral equation (6.33) now assumes the form

$$\beta\int_0^\infty e^{-\alpha|s-t|}e^{-(\beta/2)(s+t)}f(t)dt=\lambda f(s). \qquad (7.15)$$

Putting

$$\psi(t)=\exp\left(\frac{\beta}{2}t\right)f(t) \qquad (7.16)$$

we have

$$\int_0^\infty \exp(-\alpha|s-t|)\exp(-\beta t)\psi(t)dt=\frac{\lambda}{\beta}\psi(s). \qquad (7.17)$$

This integral equation is solved by[6]

$$\psi(s)=J_{2\alpha/\beta}\left[2\left(\frac{2\alpha}{\beta\lambda}\right)^{\frac{1}{2}}e^{-\beta s/2}\right] \qquad (7.18)$$

and the eigenvalues are determined by

$$J_{(2\alpha/\beta)-1}[2(2\alpha/\beta\lambda)^{\frac{1}{2}}]=0. \qquad (7.19)$$

Thus the eigenvalues of our integral equation are expressible in terms of roots of Bessel functions and are seen to be simple. It is interesting to note that if $2\alpha/\beta=\frac{3}{2}$ we can express $P(V)$ in terms of one of the theta functions. In fact, if $2\alpha/\beta=\frac{3}{2}$, Eq. (7.19) assumes the form

$$J_{\frac{1}{2}}[(6/\lambda)^{\frac{1}{2}}]=0 \qquad (7.21)$$

and since

$$J_{\frac{1}{2}}(x)=\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\frac{\sin x}{\sqrt{x}} \qquad (7.22)$$

we see that

$$\lambda_n=\frac{6}{\pi^2}\frac{1}{n^2}\quad (n=1,2,\cdots) \qquad (7.23)$$

and formula (6.36) for $V>0$ becomes

$$P(V)=\frac{\pi^2}{6}\sum_{s=1}^{\infty}\frac{s^2\exp(-\pi^2s^2V/6)}{\prod_j'(1-(s^2/j^2))}. \qquad (7.24)$$

It can be shown quite easily that

$$\prod_j'\left(1-\frac{s^2}{j^2}\right)=-\frac{\cos\pi s}{2} \qquad (7.25)$$

and therefore,

$$P(V)=\frac{\pi^2}{3}\sum_{s=1}^{\infty}(-1)^{s+1}s^2\exp(-\pi^2s^2V/6). \qquad (7.26)$$

Introducing the theta-function notation

$$\theta(q;0)=1-2q+2q^4-2q^9+\cdots \qquad (7.27)$$

we have

$$q\theta'(q;0)=-2\sum_{s=1}^{\infty}(-1)^{s+1}s^2q^{s^2} \qquad (7.28)$$

and finally

$$P(V)=-\frac{\pi^2}{6}\exp\left(\frac{\pi^2V}{6}\right)\theta'(e^{-\pi^2V/6};0). \qquad (7.29)$$

For other values of the ratio $\alpha/\beta$ we can calculate values of $P(V)$ and plot the corresponding graphs.

### 2. The Noise Power, Averaged over a Finite Time Interval

In some problems it is of interest to know the distribution of the average

$$M=\frac{1}{\tau}\int_0^\tau I^2(t)dt. \qquad (7.31)$$

The derivation of Section 4 is clearly applicable to this case if we put

$$K(u)=\begin{cases}1/\tau, & 0<u<\tau\\ 0 & \text{elsewhere.}\end{cases} \qquad (7.32)$$

The probability density is given by (4.46) replacing $V$ by $M$ except that the $\lambda$'s are now eigenvalues of the integral equation

$$\frac{1}{\tau}\int_0^\tau \rho(s-t)f(t)dt=\lambda f(s). \qquad (7.33)$$

In the simple case

$$\rho(t)=e^{-\alpha|t|} \qquad (7.34)$$

the eigenvalues can be shown to be given by the formula

$$\lambda_m=\frac{2}{g}\frac{1}{1+y_m^2}. \qquad (7.35)$$

---

[6] For the solution and discussion of the integral equation see M. L. Juncosa "An integral equation related to Bessel functions," Duke Math. J. 12, 465–471 (1945).

where $y_m$ is the $m$th (positive) root of the equation

$$\tan gy = -2y/(1-y^2) \tag{7.36}$$

and $g = \alpha\tau$.

## VIII. APPROXIMATIONS FOR $P(V)$

At the end of Section 5, it was shown that, for infinitely wide video, $P(V)$ becomes identical with the well-known probability density of the output voltage of a quadratic detector. We will derive now an approximation for the case that the video band width is very large compared with the i-f band width.

It can be shown that in this case one of the eigenvalues say $\lambda_1$, tends to unity while all others approach zero. In order to obtain the dependence of the eigenvalues on the band width ratio we shall write

$$A(f') = \frac{1}{B}a\left(\frac{f'}{B}\right) \tag{8.11}$$

and

$$C(f) = c\left(\frac{f-f_0}{b}\right), \tag{8.12}$$

where

$$\int_{-\infty}^{\infty} a(\nu)d\nu = 1 \tag{8.13}$$

and

$$c(f_0/b) = 1 \tag{8.14}$$

in conformity with our original normalization for $A(f)$ and $K(t)$. The limiting process is carried through for $B/b = \epsilon \rightarrow 0$, while the functions $a(\nu)$ and $c(\nu)$ remain unchanged. In terms of $A$ and $C$ our integral equation becomes

$$\int_{-\infty}^{\infty} A(f_1)C(f_2-f_1)g(f_2)df_2 = \lambda g(f_1) \tag{8.21}$$

and traces $S_k$ are, therefore,

$S_1 = 1,$

$$S_2 = \int\int_{-\infty}^{\infty} d\nu_1 d\nu_2 a(\nu_1)a(\nu_2) \times c(\epsilon[\nu_2-\nu_1]-f_0/b)c(\epsilon[\nu_1-\nu_2]-f_0/b), \tag{8.22}$$

$$S_3 = \int\int_{-\infty}^{\infty}\int d\nu_1 d\nu_2 d\nu_3 a(\nu_1)a(\nu_2)a(\nu_3) \times c(\epsilon[\nu_2-\nu_1]-f_0/b)c(\epsilon[\nu_1-\nu_3]-f_0/b) \times c(\epsilon[\nu_3-\nu_2]-f_0/b),$$

etc. Expanding $c(\epsilon\nu - f_0/b)$ in a power series around $f_0/b$ we shall write

$$c(\epsilon\nu - f_0/b) = 1 + C_1\epsilon\nu + C_2\epsilon^2\nu^2 + \cdots. \tag{8.23}$$

Assuming that

$$\int_{-\infty}^{\infty} a(\nu)\nu^4 d\nu$$

is finite we obtain the following approximations:

$$\begin{aligned} S_2 &\cong 1 + \alpha\epsilon^2 + 0(\epsilon^4), \\ S_3 &\cong 1 + 3\alpha\epsilon^2/2 + 0(\epsilon^4), \\ S_4 &\cong 1 + 2\alpha\epsilon^2 + 0(\epsilon^4), \end{aligned} \tag{8.24}$$

where

$$\alpha = 2(2C_2 - C_1^2)\int_{-\infty}^{\infty} a(\nu)\nu^2 d\nu \tag{8.25}$$

and essential use has been made of the fact that $a(\nu) = a(-\nu)$. Noticing that

$$\lambda_1^2 S_2 > S_4 \tag{8.26}$$

(since $S_n = \sum_1^\infty \lambda_k^n$, and $\lambda_1$ is by defintion the largest eigenvalue) we obtain

$$\lambda_1^2 > S_4/S_2 = 1 + \alpha\epsilon^2 + 0(\epsilon^4). \tag{8.27}$$

Therefore, $\sum_2^\infty \lambda_k^2$ is of order $\epsilon^4$ or higher, and $\sum_2^\infty \lambda_k^4$ is of order $\epsilon^8$ or higher. Now we notice that

$$(S_4)^{\frac{1}{4}} = \lambda_1 + 0(\epsilon^8) \tag{8.28}$$

so that

$$\lambda_1 = 1 + \alpha\epsilon^2/2 + 0(\epsilon^4). \tag{8.29}$$

Making use of the expression

$$V = \sum^{\infty} \lambda_k Z_k \tag{8.31}$$

[cf. Eqs. (4.42), (4.43)], we write

$$V = V_1 + v \tag{8.32}$$

with

$$V_1 = 1 - \lambda_1 + \lambda_1 Z_1 \tag{8.33}$$

and

$$v = \sum_2^\infty \lambda_k(Z_k - 1). \tag{8.34}$$

If $P_1(V_1)$ is the probability density of $V_1$, given by

$$P_1(V_1) = 0 \quad \text{for} \quad V_1 < 1 - \lambda_1 \tag{8.35}$$

and

$$P_1(V_1) = \lambda_1^{-1} \exp\frac{V_1 - (1-\lambda_1)}{\lambda_1} \quad \text{for} \quad V_1 > (1-\lambda_1),$$

and $w(v)$ is the probability density of $v$, we have for $P(V)$ the exact expression

$$P(V)=\int_{-\infty}^{\infty} dvP_1(V-v)w(v)$$

$$=\int_{-\infty}^{V-1+\lambda_1} dvP_1(V-v)w(v). \quad (8.36)$$

Since the width of $w(v)$ is of order $\epsilon^2$ we can use $P_1(V)$ as the first approximation of $P(V)$. To discuss the reliability of this appproximation we expand $P(V)$ in the form

$$P(V)=P_1(V)-P_1'(V)\int_{-\infty}^{V-1+\lambda_1} vdvw(v)$$

$$+\tfrac{1}{2}P_1''(V)\int_{-\infty}^{V-1+\lambda_1} v^2dvw(v)$$

$$+\frac{1}{3!}\int_{-\infty}^{V-1+\lambda_1} v^3dvw(v)P_1'''(V-\xi) \quad (8.37)$$

where, in the last term,

$$0<\xi<|v|.$$

For any fixed value of $V>0$, the integrals in the first and second term can be replaced by $\bar{v}=0$ and $\langle v^2\rangle=\sum_2^\infty\lambda_k^2$ with an error of order $\epsilon^4$ or smaller, because

$$\int_{V-(1-\lambda_1)}^{\infty} vdvw(v)$$

$$\leq\int_0^\infty \frac{v^4}{[V-(1-\lambda_1)]^3}w(v)dv=0(\epsilon^8), \quad (8.38)$$

$$\int_{V-(1-\lambda_1)}^{\infty} v^2dvw(v)$$

$$\leq\int_0^\infty \frac{v^4}{[V-(1-\lambda_1)]^2}w(v)dv=0(\epsilon^8).$$

The last term in (8.37) can be shown to be of order $\epsilon^6$ or higher. Thus we have

$$P(V)=P_1(V)+\tfrac{1}{2}P_1''(V)\langle v^2\rangle_{\mathrm{Av}}+0(\epsilon^6) \quad (8.39)$$

for any fixed $V>0$. Substituting

$$\lambda_1=1+\alpha\epsilon^2/2+0(\epsilon^4)$$

and

$$\langle v^2\rangle_{\mathrm{Av}}=\sum_{k=2}^{\infty}\lambda_k^2=0(\epsilon^4),$$

we see that for fixed $V>0$, $P_1(V)$ is a good approximation except for terms of order $\epsilon^4$ and higher.

If, in addition, the existence of

$$\int_{-\infty}^{\infty} a(\nu)\nu^6d\nu \quad (8.41)$$

is assumed the approximation can be improved by including the term

$$\tfrac{1}{2}P_1''(V)\langle v^2\rangle_{\mathrm{Av}}$$

where

$$\langle v^2\rangle_{\mathrm{Av}}=S_2-(S_6)^{\frac{1}{3}}+0(\epsilon^{12}). \quad (8.42)$$

and $\lambda_1$ [which occurs in $P_1(V)$] is calculated from

$$\lambda_1=(S_4)^{\frac{1}{4}}+0(\epsilon^8). \quad (8.43)$$

Substituting these values we obtain an approximation for $P(V)$ valid up to terms of order $\epsilon^4$ inclusive. In the neighborhood of $V=0$, no approximation which utilizes only the traces of a limited number of iterated kernels, can be expected to hold, since there—as can be seen from the general solution—the eigenvalues individually determine the shape.

The considerations of this section cannot be expected to be applicable to the case treated exactly in Section 7 since there not even

$$\int_{-\infty}^{\infty} a(\nu)\nu^2d\nu$$

exists. In that case one sees directly that the eigenvalues are such that

$$\lambda_1=1-0(\epsilon), \quad \sum_2^\infty\lambda_k^2=0(\epsilon^2).$$

In order to be able to obtain a general approximation for wide videos one must demand that the i-f spectrum falls off sufficiently rapidly at infinity.

In the opposite limit, $b/B\to 0$, similar considerations show that $P(V)$ becomes Gaussian with mean $\langle V\rangle_{\mathrm{Av}}=1$ and spread proportional to $(b/B)^{\frac{1}{2}}$.

## IX. APPLICATION TO A DISCRETE CASE

For some applications of noise theory it is desirable to have the characteristic function

$$\exp\left\langle i\sum_{k=1}^{n}\xi_kI_k\right\rangle_{\mathrm{Av}}=\phi(\xi_1\cdots\xi_n) \quad (9.11)$$

of the joint probability of $n$ successive values $I_1, I_2\cdots, I_n$; $I_k=x^2(t_k)+y^2(t_k)$ of the detector output, from which the averages of products $\langle\prod_k I_k^{m_k}\rangle_{\mathrm{Av}}$ ($m_k$ integer) can be derived. $\phi(\xi_1\cdots\xi_n)$ can be obtained by carrying through the idea of transformation to principal axes for the discrete case, that is, by a slight modification of Section 6. However, we shall derive $\phi(\xi_1\cdots\xi_n)$ from the results of Section 4 since this seems the quicker way.

We simply chose for $K(t)$ the function

$$K(t)=\frac{1}{\xi}\sum_1^n \xi_k\delta(t-t_k) \qquad (9.12)$$

where $\xi=\sum_1^n\xi_k$ for the sake of normalization, and the time variable $t$ is chosen in such a way that all $t_k<0$. For this choice of $K(t)$ we have

$$V=\int_0^\infty K(t)[x^2(t)+y^2(t)]dt=\frac{1}{\xi}\sum_1^n \xi_k I_k \qquad (9.13)$$

and, provided that the $I_k$ are normalized so that $\langle I_k\rangle_{\mathrm{Av}}=1$, we have

$$\phi(\xi_1\cdots\xi_n)\equiv e^{i\sum\xi_k I_k}=\langle e^{i\xi V}\rangle_{\mathrm{Av}}=\frac{1}{\prod_j(1-i\lambda_j\xi)}. \qquad (9.14)$$

The integral equation, whose eigenvalues are $\lambda_j$ degenerates in this case into a system of linear equations for $n$ discrete variables $f(t_k)=f_k$:

$$\frac{1}{\xi}\sum_{k=1}^n \xi_k\rho_{km}f_k=\lambda f_m, \qquad (9.15)$$

where

$$\rho_{km}=\rho(t_k-t_m)$$

with

$$\Lambda=\xi\lambda.$$

We write these equations as

$$\sum_{k=1}^n \xi_k\rho_{km}f_k=\Lambda f_m. \qquad (9.16)$$

To obtain $\phi$, it is not necessary actually to determine the eigenvalues since the product $\prod_j(1-i\Lambda_j)$ can be expressed in terms of the determinant det $|\xi_k\rho_{km}+i\delta_{km}|$ as follows:
We note that

$\prod_1^n(iu-i\Lambda_j)$ is a polynomial in $u$ which has zeros at $u=\Lambda_j$ and starts with $i^n u^n$. The determinant det$|\xi_k\rho_{km}-u\delta_{km}|$ is a polynomial in $u$, which has the same roots and starts with $(-)^n u^n$. We therefore have

$$\prod_1^n(iu-i\lambda_j)=i^{-n}\det|\xi_k\rho_{km}-u\delta_{km}| \qquad (9.17)$$

and, specially, for $u=-i$

$$\prod_1^n(1-i\Lambda_j)=i^{-n}\det|\xi_k\rho_{km}+i\delta_{km}|.$$ [7]

The characteristic function is thus obtained as

$$\langle e^{i\sum\xi_k I_k}\rangle_{\mathrm{Av}}=\frac{i^n}{\det|\xi_k\rho_{km}+i\delta_{km}|}. \qquad (9.18)$$

While it has not been possible to evaluate the joint probability

$$W_n(I_1\cdots I_n)=\frac{1}{(2\pi)^n}\int\cdots\int d\xi_1\cdots d\xi_n \times\phi(\xi_1\cdots\xi_n)e^{-i\sum\xi_k I_k}, \qquad (9.19)$$

explicitly (except for $n=1$ and $n=2$, in which cases they are well known) one can obtain averages of the form

$\langle\prod_k I_k^{m_k}\rangle_{\mathrm{Av}}$ for integer values of $m_k$,

by expanding both sides of (9.18) in power series in the variables $\xi_k$ and by comparing coefficients.

---

[7] For the general kernel a corresponding relation can be derived between the above product and the Fredholm determinant of the kernel.

# Heating and Cooling of Bolometer Elements

J. N. Shive
*Bell Telephone Laboratories, Inc., Murray Hill, New Jersey*

(Received January 20, 1947)

The analogy between thermal and electrical networks is used to obtain solutions to problems relating to the heating and cooling of air-backed and solid-backed bolometer elements exposed to radiation of periodically fluctuating intensity. From the equations derived, thermal response-frequency characteristics are computed for "typical" bolometer elements. Features of these characteristics are discussed, and the physical influences of the various thermal properties of the bolometer material and of the backing material are pointed out.

## INTRODUCTION

In some of the present-day applications of radiation bolometry the heat-sensitive receiver element is placed in series with another resistance, and a d.c. biasing voltage is applied across this series. The radiation incident upon the element is mechanically interrupted, as by a rotating sector disk, so that the element alternately heats and cools. Under these conditions an a.c. voltage is generated across the element, which, when amplified and measured, serves as an indication of the intensity of the incoming radiation.

The amplitude of the a.c. signal thus generated is approximately proportional[1] to the biasing voltage across the receiver element. This, in turn, is limited by the rate at which the element can dissipate the Joule heat produced by the bias. If the element is placed in thermal contact with a solid backing through which this heat can be more rapidly dissipated, the bias voltage can be increased. The amplitude of the temperature fluctuation of the element due to the absorption of a given chopped radiation signal is reduced by the presence of such a backing, but this reduction, for certain ranges of chopping frequency, can be more than compensated for by the use of the higher biasing voltage made possible by the backing. In addition, the use of a backing brings about a faster approach to temperature equilibrium at each interruption of the radiation beam, so that the time constant of the element is correspondingly reduced, a consideration of importance, for example, in the rapid scanning of an infra-red spectrum.

There are a number of questions in connection with the backing of bolometer elements for which both experimental and theoretical answers have been sought. It is of importance to know, for example, what thermal properties and what geometry the backing should have in order to secure maximum dissipation of the bias heating with minimum reduction of the amplitude of the temperature fluctuations of the element. It is equally important to know the effect of such a backing on the frequency dependence of the element.

It is the purpose of this paper to demonstrate a method for obtaining answers to such questions by applying the mathematics of electrical network theory. The electrical analogy has long been useful in the solution of problems in mechanics and acoustics, and some applications have already been made in the field of heat flow.[1–4] Analytical expressions governing the behavior of bolometers

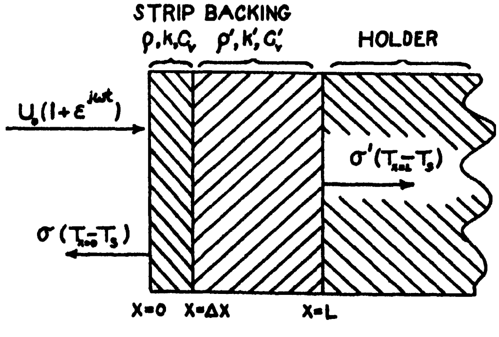


Fig. 1. Physical arrangement of bolometer strip, backing, and holder.

---

[1] V. Paschkis and M. P. Heisler, J. App. Phys. 17, 246 (1946).

[2] C. D. McCann and H. E. Criner, Machine Design 17, 137 (1945).

[3] V. Paschkis and H. D. Baker, Trans. A.S.M.E. 64, 105 (1942).

[4] M. Avrami and V. Paschkis, Trans. A.I.C.E. 38, 631 (1942).

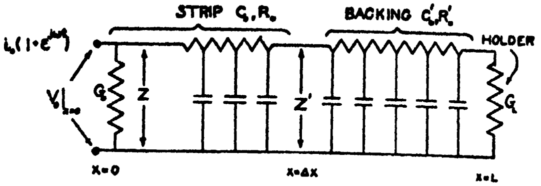


FIG. 2. Electrical analog of the bolometer assembly of Fig. 1.

will be developed from this analogy, and these will be employed in presenting graphical illustrations of the temperature fluctuation of the element in its dependence on the thermal properties of the bolometer and the backing, and on the chopping frequency.

### General Equations for Backed Bolometers

Consider a thin strip of temperature-sensitive bolometer material of 1 sq. cm area and $\Delta x$ thickness. Consider this strip to be fastened to a solid slab of backing material, also of 1 sq. cm area, and extending in thickness from $x=\Delta x$ to $x=L$. At $x=L$ let the backing slab be fastened in turn to some kind of holder or housing. This physical arrangement is sketched in Fig. 1.

Let the exposed front surface of the bolometer strip absorb an incident radiation signal of

$$U_0(1+\epsilon^{j\omega t})^* \tag{1}$$

watts, where $U_0$ is half the radiant power per unit area in the unchopped beam, $\omega$ is $2\pi$ times the chopping frequency, and $t$ is the time. Let heat be dissipated by reradiation and convection from the front face of the strip, according to Newton's law of cooling, at a rate of $\sigma(T_{x=0}-T_a)$ watts, where $\sigma$ is the dissipation constant of the front face in watts per sq. cm per degree, $T_{x=0}$ is the instantaneous temperature of the front face, and $T_a$ is the ambient temperature. Similarly let heat be lost from the rear surface of the backing slab to the holder at a rate $\sigma'(T_{x=L}-T_a)$ watts, where $\sigma'$ is the backing-to-holder dissipation constant (here considered to be thermally nonreactive) and $T_{x=L}$ is the instantaneous temperature of the rear face of the backing. These various heat flows are indicated in Fig. 1. Consider the lateral surfaces of the bolometer and backing to be thermally insulated so that heat can flow only in the $x$ coordinate.

Let $k$ and $k'$ be the thermal conductivities of the bolometer material and backing material, respectively, in watts per sq. cm per degree per cm, $\rho$ and $\rho'$ the respective densities in g per cc, and $C_v$ and $C_v'$ the respective specific heats in Joules per g degree. Then the respective thermal diffusivities $K$ and $K'$ are constants given by

$$K=k/\rho C_v \quad \text{and} \quad K'=k'/\rho' C_v'.$$

The electrical analog of the backed bolometer strip is a short transmission line (Fig. 2) having distributed series resistance $R_0$ and shunt capacity $C_0$ per unit length, looking into a longer transmission line representing the backing and having distributed series resistance $R_0'$ and shunt capacity $C_0'$ per unit length. The front-surface dissipation of the bolometer strip corresponds to a conductance $G_0$ hung across the input end of the short line, while the backing-to-holder dissipation corresponds to a conductance $G_L$ terminating the long line. The following table of equivalents may be found useful in visualizing the correspondence of analogous thermal and electrical entities (Table I).

The problem is now that of finding the input voltage amplitude $V_0|_{x=0}$ as a function of frequency when a current

$$i_0(1+\epsilon^{j\omega t})$$

is fed into the network. The a.c. solution of this network may be obtained by considering the problem as that of a short line of input impedance $Z$ (taken just inside the input dissipation conductance) and terminated at $x=\Delta x$ by the input

TABLE I.

| Entity | Thermal | Electrical |
|---|---|---|
| Input flow, a.c. part | $U_0\epsilon^{j\omega t}$ | $i_0\epsilon^{j\omega t}$ |
| Input amplitude | $T_0\|_{x=0}$ | $V_0\|_{x=0}$ |
| Series resistance per unit length | | |
| bolometer | $\frac{1}{k}$ | $R_0$ |
| backing | $\frac{1}{k'}$ | $R_0'$ |
| Shunt capacity per unit length | | |
| bolometer | $\rho C_v$ | $C_0$ |
| backing | $\rho' C_v'$ | $C_0'$ |
| Front-surface dissipation | $\sigma(T_{x=0}-T_a)$ | $G_0 V\|_{x=0}$ |
| Backing-to-holder dissipation | $\sigma'(T_{x=L}-T_a)$ | $G_L V\|_{x=L}$ |

---

[*] In (1) there is considered, for simplicity's sake, only the fundamental component of the multiply-harmonic radiant power input obtained by the usual mechanical chopping of the incident beam.

impedance $Z'$ of the long line, which in turn is terminated at $x=L$ by the dissipation $G_L$. Thus, the solution is

$$V_0|_{x=0}=i_0/(G_0+1/Z)$$

where $Z$ is given[5] by

$$Z=Z_0\frac{Z_0\sinh\gamma\Delta x+Z'\cosh\gamma\Delta x}{Z_0\cosh\gamma\Delta x+Z'\sinh\gamma\Delta x},$$

and where, in turn, $Z'$ is given by

$$Z'=Z_0'\frac{Z_0'\sinh\gamma'(L-\Delta x)+\dfrac{1}{G_L}\cosh\gamma'(L-\Delta x)}{Z_0'\cosh\gamma'(L-\Delta x)+\dfrac{1}{G_L}\sinh\gamma'(L-\Delta x)}. \tag{2}$$

In these equations $Z_0$ and $\gamma$ are the characteristic impedance and propagation constant, respectively, of the short line representing the bolometer strip, and $Z_0'$ and $\gamma'$ are the corresponding constants for the long line representing the backing. For series-resistance shunt-capacity lines these quantities have the values:

$$\left.\begin{aligned} Z_0&=(R_0/jC_0\omega)^{\frac{1}{2}}\\ \gamma&=(jR_0C_0\omega)^{\frac{1}{2}}\\ Z_0'&=(R_0'/jC_0'\omega)^{\frac{1}{2}}\\ \gamma'&=(jR_0'C_0'\omega)^{\frac{1}{2}} \end{aligned}\right\}. \tag{3}$$

With the quantities (3) replaced into (2), and with the corresponding thermal equivalents substituted into the resulting expressions, the amplitude of the temperature fluctuation at the front surface of the bolometer strip becomes

$$T_0|_{x=0}=U_0/(\sigma+1/Z_{\text{th}}),$$

where

$$Z_{\text{th}}=\frac{1}{k}\left(\frac{K}{j\omega}\right)^{\frac{1}{2}}\frac{\dfrac{1}{k}\left(\dfrac{K}{j\omega}\right)^{\frac{1}{2}}\sinh\left(\dfrac{j\omega}{K}\right)^{\frac{1}{2}}\Delta x+Z'_{\text{th}}\cosh\left(\dfrac{j\omega}{K}\right)^{\frac{1}{2}}\Delta x}{\dfrac{1}{k}\left(\dfrac{K}{j\omega}\right)^{\frac{1}{2}}\cosh\left(\dfrac{j\omega}{K}\right)^{\frac{1}{2}}\Delta x+Z'_{\text{th}}\sinh\left(\dfrac{j\omega}{K}\right)^{\frac{1}{2}}\Delta x},$$

and where

$$Z'_{\text{th}}=\frac{1}{k'}\left(\frac{K'}{j\omega}\right)^{\frac{1}{2}}\frac{\dfrac{1}{k'}\left(\dfrac{K'}{j\omega}\right)^{\frac{1}{2}}\sinh\left(\dfrac{j\omega}{K'}\right)^{\frac{1}{2}}(L-\Delta x)+\dfrac{1}{\sigma'}\cosh\left(\dfrac{j\omega}{K'}\right)^{\frac{1}{2}}(L-\Delta x)}{\dfrac{1}{k'}\left(\dfrac{K'}{j\omega}\right)^{\frac{1}{2}}\cosh\left(\dfrac{j\omega}{K'}\right)^{\frac{1}{2}}(L-\Delta x)+\dfrac{1}{\sigma'}\sinh\left(\dfrac{j\omega}{K'}\right)^{\frac{1}{2}}(L-\Delta x)}. \tag{4}$$

---

[5] A derivation of this expression for the input impedance of a distributed-constant line terminated by another impedance can be found, for instance, in Slater's *Microwave Transmission*, Chap. I.

The subscript th indicates that the impedances designated are thermal impedances.

### Application to Air-Backed Bolometers

The above equations can equally well be applied to air-backed bolometers. For an air-backed element the only heat loss is by air dissipation at the front and rear surfaces. The thermal impedance of the backing in this case is replaced by the appropriate air-dissipation factor, so that $Z'_{\text{th}}$ in (4) becomes simply $Z'_{\text{th}}=1/\sigma$. With this substitution the first two equations of (4) combine to give

$$T_0|_{x=0}=\frac{U_0}{\sigma+k\left(\dfrac{j\omega}{K}\right)^{\frac{1}{2}}\dfrac{\dfrac{1}{k}\left(\dfrac{K}{j\omega}\right)^{\frac{1}{2}}\cosh\left(\dfrac{j\omega}{K}\right)^{\frac{1}{2}}\Delta x+\dfrac{1}{\sigma}\sinh\left(\dfrac{j\omega}{K}\right)^{\frac{1}{2}}\Delta x}{\dfrac{1}{k}\left(\dfrac{K}{j\omega}\right)^{\frac{1}{2}}\sinh\left(\dfrac{j\omega}{K}\right)^{\frac{1}{2}}\Delta x+\dfrac{1}{\sigma}\cosh\left(\dfrac{j\omega}{K}\right)^{\frac{1}{2}}\Delta x}}. \tag{5}$$

Two approximations to (5) can be made, each being valid over a particular range of parameters, which simplify the task of interpreting physically the results obtained from (5). For small arguments the sinh is nearly equal to the argument and the cosh is nearly equal to unity. Thus, for

$$|(j\omega/K)^{\frac{1}{2}}\Delta x|\lesssim 0.3$$

Eq. (5) becomes

$$T_0|_{x=0}\simeq\frac{U_0(\sigma\Delta x+k)}{\sigma(\sigma\Delta x+2k)+\dfrac{jk^2\omega\Delta x}{K}}. \tag{5a}$$

For large arguments the sinh and cosh are nearly equal. Hence, if

$$|(j\omega/K)^{\frac{1}{2}}\Delta x|\gtrsim 1.8$$

Eq. (5) becomes

$$T_0|_{x=0}\simeq\frac{U_0}{\sigma+k(j\omega/K)^{\frac{1}{2}}}. \tag{5b}$$

Equations (5), (5a), and (5b) can be used to calculate theoretical response-frequency curves for various bolometers. By thermal response is meant the ratio $T_0|_{x=0}/U_0$ of the amplitude of the front-face temperature fluctuation to the amplitude of the incoming chopped radiation signal. Such a response-frequency curve is given by the solid curve of Fig. 3 for a typical bolometer strip made of thermistor material* and having:

$\Delta x=0.0018$ cm
$k=0.01$ watt/cm deg.
$\rho=5$ g/cc
$C_v=0.75$ joule/g deg.
$\sigma=0.005$ watt/sq. cm deg.

This curve is composed of the following three distinct regions, each being characteristic of a particular frequency range.

I. A frequency-independent region for very low frequencies in which, physically, the temperature of the entire flake is always in equilibrium with the incident radiation. From Eq.

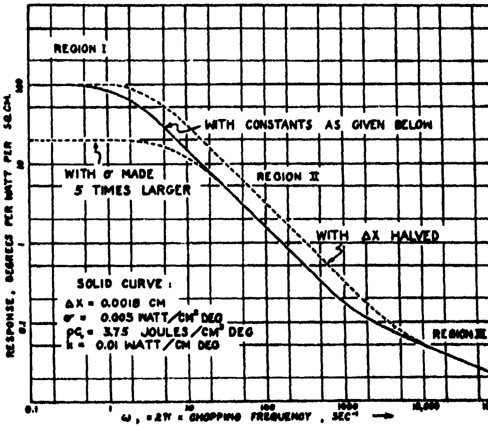


FIG. 3. Response-frequency characteristic for air-backed bolometer.

---

* Thermistor material is a mixture of certain semiconductor oxides. It has a large negative temperature coefficient of resistance.

(5a), appropriate to this region, it is apparent that the temperature fluctuations of the strip are in phase with the fluctuations of the incoming radiation.

II. An intermediate region in which the response is inversely proportional to the frequency. Here the strip temperature is no longer in equilibrium with the incoming radiation, and the phase of the former lags that of the latter by a quarter cycle. As will be shown below, the instantaneous temperature of the whole strip is effectively equal to that of the front surface over the entire frequency region for which this relationship holds. The thermal time constant of the strip is the reciprocal of the frequency at which the transition from region I to region II occurs.

III. A high frequency region in which the response obeys a $1/\sqrt{\omega}$ relationship, as given by (5b). Here the phase of the front-face temperature lags that of the incoming radiation by an eighth cycle.

The physical meaning of the transition from region II to region III can be understood from a consideration of the speed $v$ and wave-length $\lambda$ with which waves of temperature traverse the thickness of the strip in consequence of the periodic heating and cooling of the front face.

The propagation constant $\gamma$ of the electrical line representing the bolometer, as given by Eq. (3) is: $\gamma=(jR_0C_0\omega)^{\frac{1}{2}}$. The imaginary part of this quantity, called the phase constant $\beta$, is: $\beta=(R_0C_0\omega/2)^{\frac{1}{2}}$. And the velocity $v$ of electrical waves along this network is given by $v=\omega/\beta=(2\omega/R_0C_0)^{\frac{1}{2}}$. The thermal analog of this velocity expression is $v_{\text{therm}}=(2K\omega)^{\frac{1}{2}}$, while the wave-length $\lambda$ of the temperature waves is

$$\lambda=\frac{2\pi v}{\omega}=2\pi(2K/\omega)^{\frac{1}{2}}.^{*} \qquad (6)$$

Comparing (6) with the condition in (5b) for which the transition from region II to region III becomes practically complete, it is evident that $\lambda_{\text{III}}\lesssim 4\Delta x$. Thus, the transition corresponds, with increasing frequency, to the physical situation in which the temperature waves have become so short that their wave-length is comparable to the thickness of the strip, and the strip is no longer at the same instantaneous temperature throughout its entire thickness. Hence for chopping frequencies lying in region III various layers within the bolometer strip have temperatures different from $T|_{x=0}$, and the electrical output response of the bolometer bridge is no longer determined by $T_0|_{x=0}$, but rather by the average throughout the thickness of the strip of the local temperature fluctuation amplitude, with due regard for phase. It can be shown* that this electrical response continues to decrease like $1/\omega$ with increasing frequency, even though the front-face temperature fluctuation amplitude falls off like $1/\sqrt{\omega}$. It appears from these considerations that for chopping frequencies in regions I and II this solution could equally well have been obtained from a lumped-constant analogy as from a transmission line analogy.

In Fig. 3 are plotted two dashed curves showing how the response-frequency characteristic is altered if a change is made in the thickness of the strip or in the value of the dissipation constant. The dashed curve characteristic of the 5-times-greater dissipation constant becomes identical with the solid curve for $\omega\gtrsim 30$. It can be seen from this illustration that if, by backing or otherwise, the dissipation from the element can be increased, an advantage will be gained in the operation of the bolometer. For, with this 5-fold greater dissipation a $\sqrt{5}$-fold greater d.c. bias can be placed across the bolometer bridge, and for $\omega\gtrsim 30$ a $\sqrt{5}$-fold increase in the electrical signal to the amplifier will be obtained for a given radiation signal. Furthermore, the time constant has been decreased to $1/\sqrt{5}$ of its former value.

Unfortunately, there is no solid backing material whose dissipation is entirely conductive, and some thermal reactance due to the volume heat capacity of the material is inescapable. The effect of this reactance, as will be shown in the next section, is to modify the advantage gained by the use of a solid backing.

## Application to Solid-Backed Bolometers

As an example of the kind of response-frequency curve to be expected for a solid-backed

---

* These expressions for wave-length and velocity of propagation of temperature waves in a semi-infinite medium were derived (otherwise) and used by Kelvin in his studies of the penetration of diurnal and annual waves of temperature into the earth.

* The author is particularly indebted to W. H. Brattain for helpful discussion of this subject.

bolometer there is shown in the solid curve of Fig. 4 the characteristic of the same strip bolometer as was used in the calculations for Fig. 3, but now fastened intimately to a rocksalt backing having the following dimensions and properties.

$$L=0.1\text{ cm}:\text{area}=1\text{ sq. cm}$$
$$k'=0.07\text{ watt/cm degree}$$
$$\rho'C_v'=2\text{ joules/cm}^3\text{ degree}$$
$$\sigma'=0.1\text{ watt/cm}^2\text{ degree.}$$

In the same figure appear a number of dashed curves showing the effects produced by varying, one or two at a time, some of the parameters of the backing. The caption adjacent to each dashed curve explains which parameter has been changed, and by how much. All of these curves have been computed from (4), using, wherever appropriate, the type of simplifications indicated in (5a) and (5b).

These curves exhibit features which are capable of simple physical explanation. The peculiar shape of the solid-line characteristic *A* is caused by the effect of the thermal reactance associated with the volume heat capacity of the backing. If this reactance could be eliminated, as for curve *B* which, to illustrate an extreme case, has been computed for $\rho'C_v'=0$, the characteristic would resemble that of an air-backed unit with unachievably high surface dissipation. This circumstance is of importance in the practical selection of a backing material. It can be concluded generally that for backings having low $\rho'C_v'$ the response can be expected to remain frequency-independent out to higher chopping frequencies than for backings having comparable $k'$ but higher $\rho'C_v'$.

Curve *C*, which has been computed for the extreme case $k'=0$, appears to coincide exactly with the solid curve of Fig. 3 except in the frequency-independent region, where the limiting response is twice that of the air-backed unit. With $k'=0$, the rear surface of the bolometer strip is thermally insulated, so that it now has only half the total dissipation of the air-backed unit. Of course, such a low-conductivity backing would be of no practical value in dissipating the bias heat of the element.

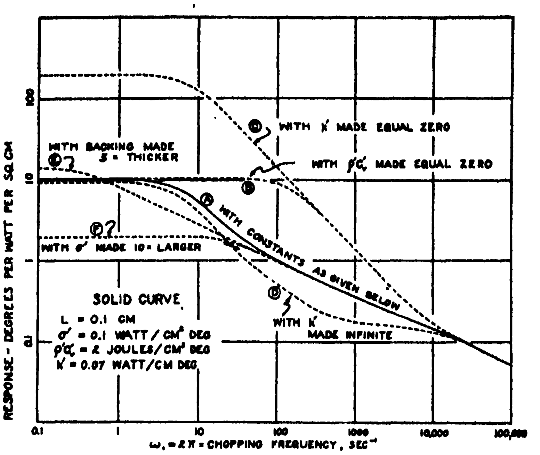


FIG. 4. Response-frequency characteristic for rocksalt-backed bolometer.

In curve *D*, to illustrate still another extreme case, the backing has been given an infinite thermal conductivity. This is equivalent to placing the entire thermal capacity of the backing directly in contact with the rear surface of the bolometer strip. This unbuffered reactance brings about the rapid decrease in response in the range from $\omega=10$ to $\omega=100$ and results in the nearly frequency-independent response in the range from $\omega=500$ to $\omega=10{,}000$. In the latter range the frequency is so high that the heat capacity of the backing presents practically a "short-circuiting" reactance to the rear surface of the bolometer strip. Thus, while the temperature at the front surface of the strip is being driven by the fluctuation of the incoming radiation, the temperature of the rear surface is being held nearly constant by this reactance. And, while the whole strip is at the same phase throughout, it is not at the same temperature throughout. This state of affairs does not change much with increasing frequency until so high a frequency is reached that the wave-length of the thermal waves in the strip becomes shorter than the thickness of the strip. Then, of course, the front-surface temperature of the strip becomes independent of the nature and properties of the backing.

It is to be noted that in the solid curve *A* there appears some tendency toward a levelling-off in the neighborhood of $\omega\sim 100$. The reason is as explained in the last paragraph, except that here the full short-circuiting effect of the backing reactance is buffered by the distributed thermal resistance of the backing, and the levelling-off is not as pronounced.

Curve $E$ shows the modification in the response-frequency curve brought about by making the backing five times thicker than for the solid curve $A$. Increasing the thickness of the backing increases the heat capacity, and the parallel reactance of this extra capacity makes itself felt at a lower chopping frequency than for curve $A$. Curve $E$ tends toward a higher low frequency limiting response than curve $A$ because of the decrease in dissipation, brought about by the increased thermal resistance of the thicker backing.

Finally, curve $F$ shows the effect of making the backing-to-holder dissipation ten times larger than for curve $A$. This increased dissipation lowers the low frequency limiting response and makes it necessary to go to higher chopping frequencies before coming to the fall-off due to the heat capacity of the backing.

It is often stated that an ideal backing material is one having a large value of the diffusivity $K'$. In making such a statement a distinction must be made whether the diffusivity is desired to be large by virtue of a large thermal conductivity $k'$ or by virtue of a small volume heat capacity $\rho' C_v'$. A comparison of curves $B$ and $D$ shows why this distinction is necessary. In a bolometer assembly of such geometry and thermal properties as are considered here the highest permissible d.c. bias across the bolometer strip is determined chiefly by the backing-to-holder dissipation rather than by the thermal conductivity of the backing. As long as the latter is large enough (as it is for rocksalt) so that the backing offers a comparatively small thermal resistance between the element and the holder, there is little virtue in having $k'$ still higher, as, for example, for quartz. However, any decrease in the value of $\rho' C_v'$ improves the response in the frequency region from $\omega \sim 1$ to $\omega \sim 1000$ and also decreases the time constant, both without affecting the permissible d.c. bias.

## DISCUSSION

In the interest of simplifying the development of this presentation a number of short cuts and omissions have been employed which would have to be taken into account in the application of the method here outlined to any physical problem where accuracy of theoretical prediction is desired. There will be pointed out, in the next few paragraphs, the nature of some of these simplifications, their consequences, and, so far as possible, their specific remedies.

All the equations appearing in this paper were derived on the assumption of a single sinusoidal amplitude fluctuation of input radiant power. If the fluctuation is obtained with a sector disk producing a nearly square-wave input of radiant power, a more satisfactory solution would be obtained by solving the appropriate equation for the fundamental chopping frequency and two or three of its harmonics as well, assigning to each harmonic solution a weight corresponding to the Fourier coefficient of the particular harmonic in the square-wave expansion, and summing over the expansion.

While backed bolometers can be made for which the thermal contact between the element and its backing is intimate, it is usually necessary in practice to fasten the two together with a thin layer of cement whose thermal properties may contribute to the behavior of the whole. The backed bolometers discussed in connection with the curves of Fig. 4 have been idealized to the extent of neglecting this layer and its effects. A paper is now in preparation by W. van Roosbroeck in which some of these details are treated. Van Roosbroeck's treatment will also include a consideration of cases in which the incoming radiation, instead of being absorbed, as here assumed, at the front surface of the bolometer strip, is absorbed throughout the thickness of the strip.

In usual practice a backed bolometer is made of a narrow strip of sensitive material cemented to a backing slab much larger in area than the strip. Hence the physical problem is not one dimensional except in a small volume of the backing immediately adjacent to the strip. At larger distances away from the strip in the backing the heat flows along nearly radial lines and the isothermals are nearly semi-cylinders coaxial with the center line of the strip. The analogous electrical problem is one in which the series conductance and parallel capacitance of the transmission line representing the backing increase nearly linearly with distance along the line. Except for special cases exact solutions for varying-parameter lines are not easily obtainable. However, approximate electrical analogs can be set up which are capable of solution. Such approximations, over certain frequency ranges are assisted by the fact that temperature waves in a semi-infinite solid are attenuated in amplitude by a factor $e^{-2\pi}$ per wave-length. Hence, for higher frequencies the temperature waves, which become shorter in wave-length with increasing frequency, do not penetrate far enough into the backing to get into the region where the fanning-out of the heat flow lines occurs, and the constant-parameter solutions already derived can be used as good approximations.

In deriving the equation for the solid-backed element the assumption was made that the holder presented a non-reactive dissipation to the rear surface of the backing slab. This assumption could lead to error only in frequency region *I* where the chopping rate is so low that the temperature waves penetrate clear through the backing. For higher frequencies the exact nature of the holder's impedance would not be of any consequence.

## ACKNOWLEDGMENTS

The author is indebted for assistance in the development of this method and in the interpretation of some of its results to W. H. Brattain and J. A. Becker. The helpful advice of R. O. Grisdale in the arrangement of this manuscript is likewise acknowledged.

# The Speed of Propagation of Brittle Cracks in Steel

G. Hudson* and M. Greenfield**

*David Taylor Model Basin, USN, Washington, D. C.*

(Received September 30, 1946)

Brittle failure was produced in notched steel tensile specimens. The speed of propagation of the brittle crack was measured and found to be approximately 40,400 in./sec.

THERE have been recent reports of the brittle failure of steel plate used in a variety of engineering structures. This rupture occurred when the steel was subjected to certain severe conditions. In the main these brittle failures may be characterized by the following:

(1) A crack is propagated with extreme rapidity through the steel plate,

(2) The surface of rupture created by this crack is orthogonal to the plane of the plate,

(3) This rupture surface shows a characteristic "herringbone" pattern,

(4) There exist small permanent strains, often of less than 2 percent, in the neighborhood of the fractured surface. It was thought that a study of the speed of propagation of brittle cracks in steel might contribute to an understanding of the mechanisms of these phenomena.

A preliminary note on the beginning of this study has been given in an earlier paper.[1] This work was subsequently extended with a more refined technique, to include more variation in the path length over which the average velocity of the crack was measured, and a variation in the distance of this path from the point of initiation of the crack. In the present paper, in addition to reporting the later results which agree with the previous ones within the limit of experimental error, more details are given of the technique used.

It has been found possible to produce brittle failure in a steel tensile specimen at room temperatures in a controlled experiment of the following type. The specimens consisted of rectangular plates of a plain carbon structural steel (also known as medium steel) with a carbon content not exceeding 0.31. This steel had a minimum average ultimate strength of 60,000 lb./in.$^2$, and a minimum average yield strength of 34,000 lb./in.$^2$. This steel had not been heat treated. The specimens were approximately 2 feet long, $5\frac{3}{4}$ inches wide, and 1 inch thick. To fix

---

* Now at New York University, New York City, New York.

** Now at North American Aviation, Los Angeles, California.

[1] M. Greenfield and G. Hudson, Proc. Nat. Acad. 31, 150 (1945).

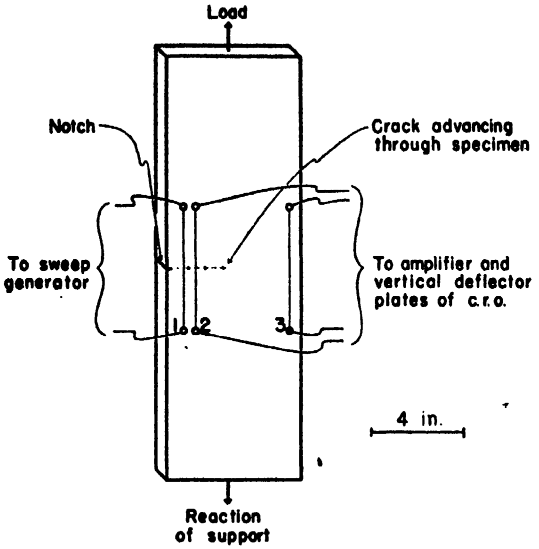

Load
Notch
Crack advancing through specimen
To sweep generator
To amplifier and vertical deflector plates of c.r.o.
1
2
3
4 in.
Reaction of support

FIG. 1. Schematic diagram showing the positions of one-mil Nichrome break wires 1, 2, and 3 on the tensile specimen.

the point of initiation of the failure, a notch consisting of a milling saw cut of the order of $\frac{1}{32}$ inch in width and $\frac{1}{4}$ inch deep was machined into one edge of the specimen midway between its ends. See Fig. 1.

The specimen was placed in the grips of a tensile machine† and pulled to failure. A brittle crack initiated at the milling cut, as indicated by Fig. 1, was propagated across the specimen, at right angles to the edge.

An attempt was made to photograph the crack as it spread through the plate. The camera used was a Fastax Western Electric, operating at about 3000 frames per second. Even this speed was insufficient to catch the phenomenon, as the crack invariably spread across the specimen between two successive frames. It was possible, however, to infer by this method a lower limit of the order of 5000 in./sec. for the velocity.

The following method using electronic recording was finally adopted for determining the speed of propagation of the crack in these specimens. The specimens were first prepared by heating them to a temperature sufficient to melt but not boil a layer of medium de Khotinsky cement. This cement was spread smoothly on the surface of the metal in the region through which the crack was to be propagated. While the de Khotinsky was still fluid, several lengths of one-mil Nichrome wire were embedded in the cement, as shown schematically in Fig. 1. These wires were arranged to be approximately parallel, about one inch apart, and perpendicular to the expected path of the crack. The de Khotinsky thus served not only to cement the wire firmly to the tensile specimen,

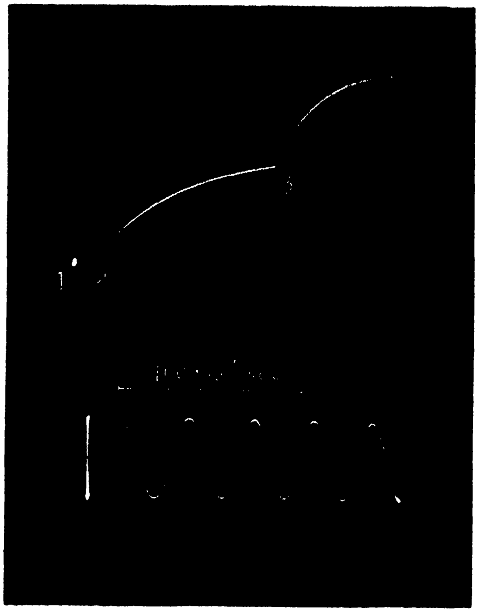


FIG. 2. A typical record obtained from the breaking of a tensile specimen like that shown in Fig. 1. The numbers 1, 2, and 3 designate the positions of the electron beam at the instants of breaking of wires 1, 2, and 3, respectively.

TABLE I. Speed of propagation of brittle crack.

| Testing machine | Specimen number | Wire numbers | Distance inches | Average velocity inches/second |
|---|---|---|---|---|
| TMB | E-1 | 2–3 | 2.23 | 42,700 |
| TMB | E-3 | 2–3 | 4.41 | 37,900 |
| NBS | F-3 | 2–3 | 1.01 | 36,500 |
|  | F-3 | 3–4 | 0.98 | 43,000 |
|  | F-3 | 4–5 | 0.96 | 39,500 |
|  | F-3 | 5–6 | 0.96 | 41,200 |
| NBS | G-2 | 3–4 | 0.96 | 42,400 |
|  | G-2 | 4–6 | 1.99 | 41,600 |
| NBS | G-3 | 2–4 | 1.97 | 44,200 |
|  | G-3 | 4–5 | 0.96 | 37,400 |
|  | G-3 | 5–6 | 0.98 | 43,700 |
| NBS | G-5 | 3–4 | 1.00 | 41,200 |
|  | C-5 | 4–5 | 0.99 | 33,000 |
|  | C-5 | 5–6 | 0.98 | 41,000 |

Average of above velocities = 40,400 ± 2500 in./sec.

---

† These specimens were actually tested in two different types of machines; one is a Baldwin Southwark hydraulic type with a 600,000 lb. capacity located at the David Taylor Model Basin, USN, Washington, D. C., and the other, an Olsen screw type with a 600,000 lb. capacity, located at the National Bureau of Standards, Washington, D. C.

but also insulated the wire from the specimen. The ends of the wires which protruded from the cement were insulated from the specimen by strips of paper stuck to the specimen partly by the de Khotinsky cement and partly by Scotch Tape. After cooling the specimen, the ends of the wires were then soldered to binding posts on Bakelite strips. These were held by rubber bands to the specimen about two inches on either side of the path of the crack.

When a crack is propagated from the base of the saw cut, it breaks the fine wires in succession. The breaking of the first wire is used to start a single horizontal sweep of the electron beam on the screen of a cathode-ray oscillograph. Each of the succeeding wires, on breaking, introduces a voltage step on the vertical deflection plates by means of a simple voltage divider arrangement. The resulting staircase-like record may be photographed, e.g., see Fig. 2. The rounding off of the successive steps is caused by the impedance of the circuit used. A calibration wave for timing is provided by a standard oscillator. See Fig. 2.

The average velocity of the crack between any two successive wires is calculated by dividing the measured distance between them by the corresponding time interval measured from the oscillograph record.

Table I gives the final results. Column 1 states where the specimen was tested, i.e., either at the Taylor Model Basin or at the National Bureau of Standards. Column 2 is the specimen number. Columns 3 and 4 give, respectively, the numbers of the 1-mil Nichrome wires between which the speed was measured, and the distance between those wires. For example, Specimen F-3 had six (6) 1-mil Nichrome wires cemented to it. The breaking of the first wire started the single sweep across the c.r.o. screen. Then there were signal wires 2 through 6 spaced as indicated in column 4. Column 5 gives the average velocity for the corresponding distance listed in column 4.

The average of all the velocities listed in column 5 is 40,400 in./sec. with a mean absolute deviation of 2500 in./sec.

It is intended to continue this work to learn whether there is a significant variation of the velocity of the brittle crack with the thickness of the steel plate, its temperature, and its composition.

---

# Use of the Broadcast Band in Geologic Mapping*

LARKIN KERWIN**
*Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Received December 4, 1946)

A review of the literature indicates that a method of geologic mapping by means of observing the effect of geologic anomalies on electromagnetic field intensity should be practical. Field equipment is designed and assembled for that purpose and proves accurate and simple to use. Preliminary investigations with it successfully locate several anomalies.

## INTRODUCTION

THE practicability of using high frequency electromagnetic waves in sub-surface geologic investigation is generally viewed as dubious. Joyce's theoretical conclusions from Maxwell's work[1] promise great attenuation under the assumed conditions. Eve,[2] Terman,[3] and Ollendorf support these considerations. However King predicted less attenuation, while Fritsch[4] declares that theoretical assumptions are never justified in practice and suggests that practical procedures are quite possible.

Relatively few experimental investigations

---

* This paper is based on a portion of a thesis entitled "The Development of Field Equipment Suitable for Measuring the Significant Properties of the Earth and Their Effect on Electromagnetic Wave Propagation," submitted for credits towards the degree of Master of Science at M.I.T. Financial support for the work was provided through a grant from The Geological Society of America, which is hereby gratefully acknowledged.

** Now with the Department of Physics, Laval University, Quebec.

[1] J. W. Joyce, Bureau of Mines Tech. Pap. No. 497 (1931).

[2] A. S. Eve and D. A. Keys, *Applied Geophysics* (Cambridge University Press, New York, 1929).

[3] F. E. Terman, *Radio Engineers' Handbook* (McGraw-Hill Book Company, Inc., New York, 1943), p. 698.

[4] V. Fritsch, Beit. Ang. Geophysik [4] 5, pp. 375-395.

have been undertaken and these are inconclusive. Potter and Friis[5] determined that ordinary topographic variations did not have great effects on high frequency field intensities. On the other hand, Cloos,[6] Howell,[7] and Fritsch[8] found that field intensities did vary with the sub-surface geology. Joyce, Eve,[9] Allen, and Silverman and Sheffet[10] worked qualitatively in mines or tunnels and found in general that broadcast frequencies often penetrate the ground for considerable distances, but that higher frequencies were quickly absorbed. Joyce also did some quantitative work at 500 cycles, but it is highly doubtful if this frequency justified the use of electromagnetic wave theory in considering the results. The effect was undoubtedly a simple inductive one in view of the tremendous wave-length. Fritsch[11, 12] lists many interesting results of successful underground propagation.

These results and others indicated that the effect of geology on the field intensity of electromagnetic waves might provide a useful method of geologic mapping. The Geologic Society of America provided a grant for the purpose of developing equipment, and preliminary results obtained with it last summer indicate that the band from 200 kc to 1000 kc will prove useful in future work.

It was decided that the investigation of the field patterns of various frequencies over the same area would provide useful information. Geologic anomalies would cause irregularities in the pattern which would vary with frequency according to the penetrating power of the various wave-lengths. Equipment was purchased and adapted or constructed to produce, monitor, and measure suitable electromagnetic fields.

## EQUIPMENT

The field intensity recorder was assembled inside a wooden-bodied station wagon supplied by the Geotechnical Corporation. Its basic component was a modified Model DAE radio direction finder. The antenna consisted of a 14-inch shielded loop, mounted in a stationary cylinder provided with azimuth scale and commutator rings connected to a shielded lead. This loop was mounted on a spring tripod which was bolted to the floor of the station wagon. This mount was firm, but protected the loop from the shocks produced by the car's motion.

The associated receiver was mounted on a small oak table, directly in front of the operator's seat. This unit was of the superheterodyne type, covering the frequency range 240 kc to 2000 kc in three bands. It was provided with volume control, antenna tuning, and beat frequency oscillator, which made the receiving of unmodulated signals considerably easier. The case was provided with shock mounts screwed to the reinforced table top. In this unit the field produced a proportionate I.F. voltage, and that of the 2nd I.F. stage was selected as the most satisfactory for recording purposes. It was tapped to a jack on the panel, the signal from which was then proportional to the field strength, but could be controlled by the volume control circuit.

This voltage was measured by a vacuum-tube voltmeter of the "Vomax" type, installed immediately above the receiver on shock mounts attached to the latter's case. Sponge rubber pads were glued to the roof of the wagon and the voltmeter as added protection. For continuous recording, an Esterline-Angus recording milliammeter was used, having speeds up to six inches per minute. It was connected into the voltmeter circuit *via* a jack mounted on the meter face. By using this system throughout, any component could be removed for alteration or replacement without necessitating a rewiring job each time. For transportation and moderate speeds of recording the meter was mounted in sponge rubber padded clamps on a shelf under the table. Here it was easily available for inking, marking the chart, etc. The needle was weighed somewhat more than usual, and it was found that it would not jump from the paper if the car was driven at low speeds. For continuous operation at high speeds, a low frequency spring suspension should be used.

---

[5] R. K. Potter and H. T. Friis, Proc. I.R.E. [4] 20 (April, 1932).

[6] E. Cloos, Am. J. Sci. [166] 28 (October, 1934).

[7] Howell, Geophys. [2] 8 (April, 1943).

[8] V. Fritsch, Beit. Ang. Geophysik [3] 5, pp. 315–364.

[9] Eve, A.I.M.E. Tech. Pub. No. 316.

[10] D. Silverman and D. Sheffet, Geophys. [4] 7 (October, 1942).

[11] V. Fritsch, Beit. Ang. Geophysik [1] 6, pp. 100–119; [3] 6, pp. 277–306; [4] 6, pp. 407–412; [4] 7, pp. 449–461.

[12] V. Fritsch, Neues Jahrb. f. Geol. B., Vol. 84, Series 1, pp. 90–116.

All of these units were operated with 110-volt, 60-cycle power. In the mobile unit this was provided by a Janette rotary converter operated by storage batteries. This produced a steady voltage at about 200 watts and operated silently. When well charged, the batteries provided 24 volts to operate the equipment for about 8 hours, more than a normal day's field work. A battery charger was mounted underneath the operator's seat and connected so that after the day's operations a long power cable could be connected to any convenient 110-volt outlet and the batteries charged overnight. This system proved completely satisfactory. A constant-voltage transformer proved useful for last-minute work at the day's end, in case the batteries began weakening. A general view of the equipment in the station wagon is given in Fig. 1.

It was intended that frequency ranges from 200 kc and up should be investigated. For preliminary work the broadcast band seemed suitable, and first field runs were made by recording the fields of various commercial transmitters as described later. For easily controllable work and for greater frequency range, a compact 75-watt transmitter was obtained. Known as the BC-375-E in the U. S. Army Signal Corps for whom it was designed, this transmitter provided continuously variable frequencies from 200 kc to 500 kc and from 1500 kc to 12,500 kc by means of six plug-in tuning units. This transmitter was designed to operate from 24-volt d.c. supply; plate voltages, etc., being supplied by a dynamotor. This supply provided approximately 750 watts. For portable operation battery operation was indicated, but was impractical because of the heavy drain of about 30 amperes. Accordingly, after some search a small gasoline driven generator was obtained which provided the required power, and delivered it through a number of "floating" storage batteries. This provided a smooth d.c. supply of about 28 volts for which the transmitter was easily modified.

During the process of taking field intensity measurements over various ground conditions, it was possible that the electromagnetic field would undergo certain variations independent of the geology. Fluctuations at the transmitter, atmospheric changes, the time of day at which readings were taken, etc. all contributed to fluctuation. It was therefore necessary to provide a stationary monitoring field intensity meter which would keep a record of these changes. The fact that relative field strengths and not absolute measurements were required simplified matters somewhat. A Hallicrafters Model SX-28 all-band receiver was obtained and provided with a simple fixed antenna. A vacuum-tube voltmeter was connected into the a.v.c. circuit of the receiver. Thus by the adjustment of the r-f amplifier amplitude and the use of a suitable voltmeter scale, a convenient deflection was obtained. This followed accurately the variations in field intensity. In order to obtain a continuous record of these changes, an Esterline Angus recording milliammeter was placed in the voltmeter circuit. This meter was adjusted to record at the rate of 1 foot per hour, which provided a satisfactory record of field strength changes without noting any minor fluctuations.

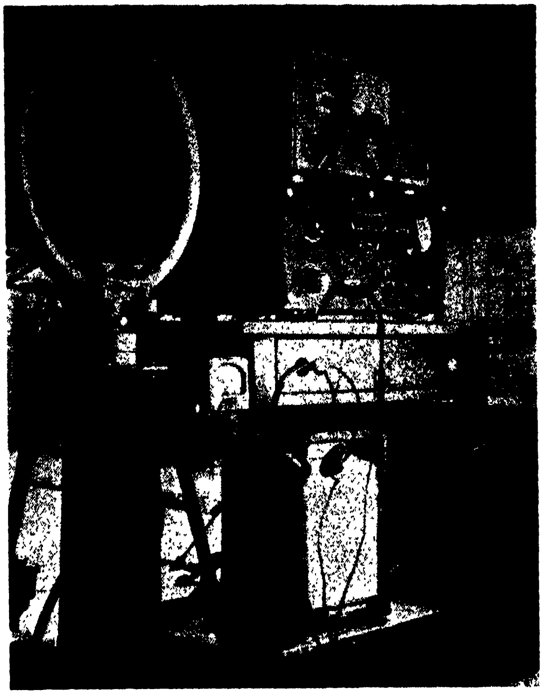
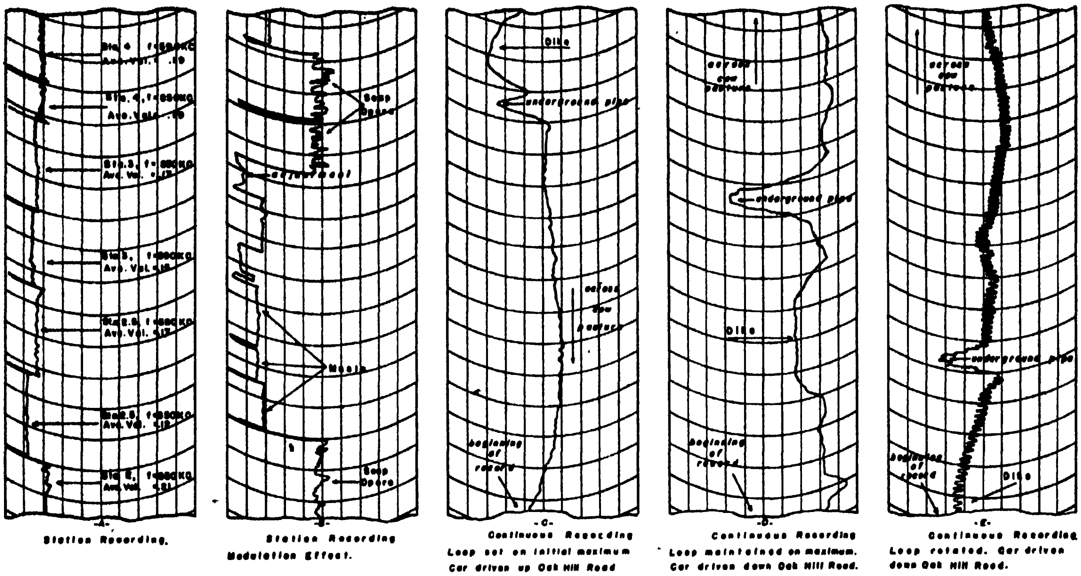


FIG. 1. Field intensity recording equipment mounted in station wagon.

All of this equipment operated from 110-volt, 60-cycle current which was usually available near the scene of operations, or could be provided by a small motor generator.

In the course of the field work which was



-A- Station Recording.

-B- Station Recording Modulation Effect.

-C- Continuous Recording Loop set on initial maximum Car driven up Oak Hill Road

-D- Continuous Recording Loop maintained on maximum. Car driven down Oak Hill Road.

-E- Continuous Recording Loop rotated. Car driven down Oak Hill Road.

FIG. 2. Field intensity recordings (taken from recording meter chart).

extended over several weeks, the station wagon was driven over all kinds of roads, and a complete gamut of speeds. The equipment and mountings were thus subjected to a most thorough testing, and after a few adjustments following the first couple of days' experience, the apparatus proved entirely field-worthy. The only precaution taken was to remove the recording

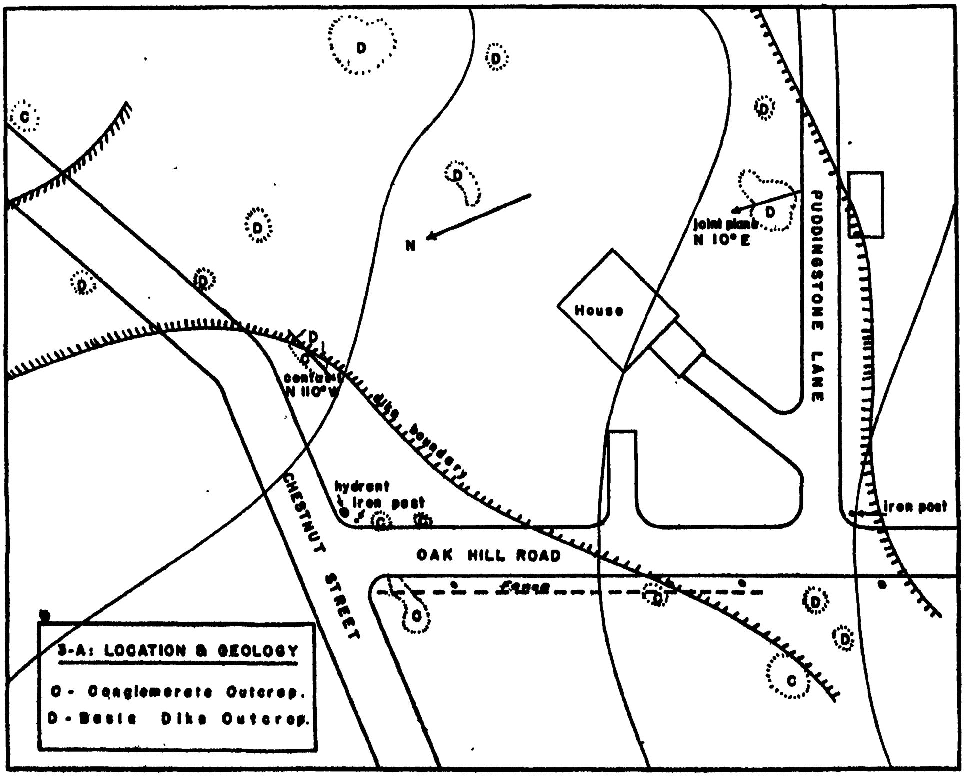


FIG. 3a. Area location and geology.

milliammeter needle if it was intended to proceed over rough ground at high speed. At low speeds it recorded satisfactorily. On this score, it is believed that future equipment patterned after this will provide a simple and rugged tool for the investigator.

### FIELD PROCEDURE

For preliminary work, a dike situated in Newton, Massachusetts was selected as a suitable testing ground. This dike had been roughly located during a survey by Mr. Blumberg of Harvard, who kindly lent some of his field maps to this project. An extension of basic melophyre was found protruding through the roxbury conglomerate which was characteristic of the area. The contrast in the physical properties of these rocks is considerable, making them quite suitable for the purpose in hand.

The district was suburban, with considerable vegetation and the houses widely separated. Some time was spent in examining the geology in order to locate the dike factually if possible. In Fig. 3a the dike location based on surface geology is indicated.

Field intensity measurements were then made under a variety of conditions. First, detailed runs were made by taking field intensity records of about five minutes duration at points about twenty-five feet apart along the road. In each case the location of the loop and its orientation were established by taping and bearings taken with a Brunton-type pocket transit. The latter readings were taken from two directions and at some distance from the station wagon, eliminating the latter's effect on the compass. The bearings so taken were all consistent with those established from surveyed maps.

The duration of the five-minute reading was important in order to achieve an averaging effect. An example of the readings obtained over several stations as taken from the field chart is given in Fig. 2a. One of the disadvantages of using the broadcast band with the present equipment here became apparent. The modulating voltage had some effect on the instantaneous readings, although not the average. Therefore readings extending over some period of time were necessary to obtain accurate values. The type of program being broadcast also has an effect, as the modulation of musical programs is much

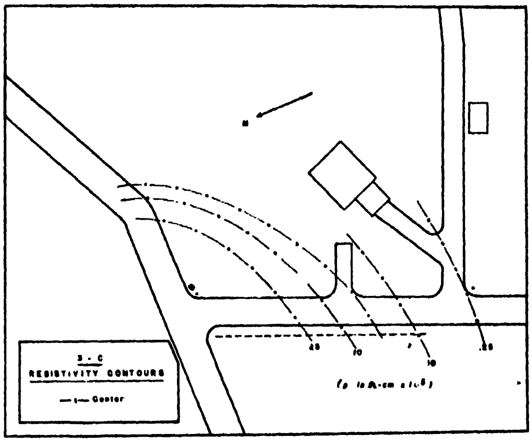


FIG. 3c. Resistivity contours.

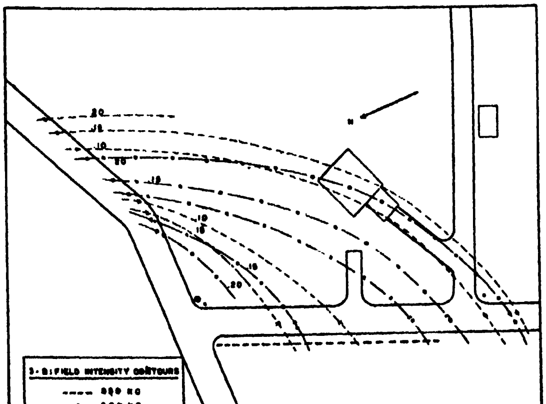


FIG. 3b. Field intensity contours.

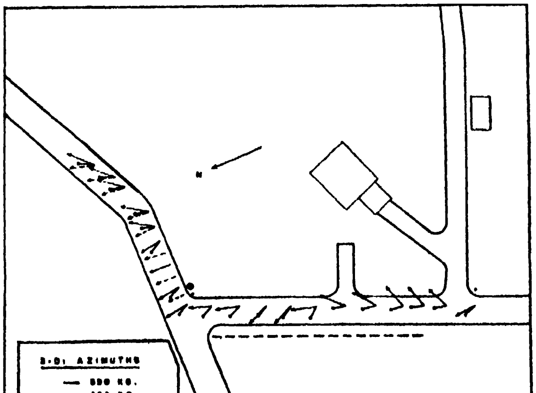


FIG. 3d. Field intensity azimuths.

smoother than those of the "soap-opera" type, and gave records which were much easier to interpret. An example is shown in Fig. 2b. These difficulties are avoided when using the unmodulated wave of the local transmitter.

The procedure was repeated several times, for several frequencies. Records were also made in the sunshine, rain, under dull clouds, and at night.

In addition to this, several types of continuous recording were investigated. By keeping the loop in a constant position of initial maximum strength and driving the station wagon slowly over the area, a figure such as that shown in Fig. 2c was recorded. This shows the prominent features which are marked. Other runs were made by keeping the loop manually oriented as much as possible in the direction of maximum intensity. The result is shown in Fig. 2d, where it is seen to bear close resemblance to the former record. Note that runs were made in varying directions and must be properly interpreted as to orientation. Continuously rotating the loop so as to produce an average effect gave a record such as shown in Fig. 2e, which follows the same general pattern as before.

A series of resistivity measurements was also made in the area in order to determine whether any relation existed between resistivity and field intensity. These measurements were made by inserting four metal stakes into the ground about twenty-five feet apart along the line of measurement. 150 volts, 60 cycle, as supplied by the small motor generator in the station wagon were applied between the two outer stakes. By measurements taken of the resistance between them the current put into the ground could be calculated. The voltage drop between the two inner stakes was then measured with a vacuum-tube voltmeter, and from these readings the resistivity was calculated.[13] By advancing the rear stake and changing terminals, successive readings were quickly taken. In this manner the resistivity was measured along the same direction as that in which the field intensity was recorded.

From the accumulated data, field intensity contours were drawn on the field map, and are shown in Fig. 3b. We see that the contours follow closely the direction of the dike as determined by the surface geology. Resistivity contours were also determined and are shown in Fig. 3c, where they assume the same general pattern as the other figures. The resistivity of the dike is substantially lower than the conglomerate, and as would be expected, the field intensity above the dike was also lower.

Azimuths of maximum field intensity are given in Fig. 3d, but they do not appear to convey any conclusive information. Over the dike, on Oak Hill Road, they seem to orient more towards the North than at other places.

Certain characteristics of the graphs drawn from the accumulated data are of some interest. It was observed that data taken at night showed considerably less variation than that taken by day. Measurements of low frequencies showed greater disturbance over the dike than did those of higher frequencies. Data taken during rainy weather showed the fields to be in general stronger and less affected by geology. Low intensities over the dike pointed to its higher conductivity, as was verified by resistivity measurements. The low frequency irregularities were more widespread than the others, indicating that the ground anomaly was more than a surface effect. To some extent here, and more so in urban areas later investigated, the effects of manholes, steel rails, etc. were quite obvious and easily detected.

Besides the detailed investigations in this area, other districts were investigated, using both stationary and rotating loop techniques. In the more urban area of Hammond Street, Brookline, a similar dibase dike protruding through the conglomerate was located by this method, and its presence verified by an examination of the surface geology. The boundaries of the dike were fairly sharply recorded as the field weakened over it. In Medford there is a noted extrusion of Medford dibase through Lynn volcanic complex. In this instance the resistivity of the dike is higher than the surrounding rocks, and the field intensity increased over it.

It was also observed that topography appeared to have relatively little effect. In driving over stretches of country in Arlington where the geology was uniform, the field remained essen-

---

[13] C. A. Heiland, *Geophysical Exploration* (Prentice-Hall, Inc., New York, 1940), p. 709.

tially constant except for manhole effects, etc., whether the receiver was passing over rolling hills, proceeding through cuttings, or on level ground. A notable exception was observed on the top of Pine Hill, a prominence beside the Medford Dike and very close to the transmitting antenna of the broadcast station being recorded. At the top of the hill the signal was extremely strong, but it weakened considerably as the bulk of the mountain was interposed sharply between the receiver and transmitter.

For most of these field measurements the field itself remained very constant, as shown by the monitoring station. Variations were observed between daytime and nighttime records, and with weather; but over considerable periods (3-4 hours) if the weather and daylight conditions were constant the field remained quite steady.

In general it is considered that the results obtained thus far show sufficient promise to warrant a continuation of the research. The equipment has proven sufficiently rugged, sensitive and simple to operate over long periods of field measurements. The cost of operation involves only the depreciation of equipment, and station wagon expenses. The method of field intensity recording has been used to locate a geologic anomaly as determined by other independent means, and has generally located other geologic features. Accordingly, further work on the subject is being done.

## MODIFICATIONS

In light of past experience, several modifications are recommended for the equipment, and will be incorporated into it shortly. The portable transmitter will provide a much more satisfactory field from the point of view of modulation as well as that of situation and pattern. For continuous recording, the milliammeter is being equipped with speedometer drive. This will permit operation by one person rather than the operator and driver previously necessary. The loop antenna is being supplemented by a vertical antenna, the two being interchangeable by means of a switch. This will eliminate the necessity of rotating the loop when azimuths are not being taken. A switching system in the charging circuit will make that operation somewhat more convenient to perform. The monitoring unit is being made self-sufficient by powering it with another rotary converter operating from 12 volts d.c. The SX-28 Hallicrafters receiver is being replaced by a smaller, more portable one.

Considerable attention is being devoted to frequencies below the broadcast band, as field experience here supports the general theory that low frequencies are more suitable. Experiments will also be performed on bands above the broadcast, up to about 12,000 kc.

## ACKNOWLEDGMENT

The writer expresses his sincere personal appreciation to Dr. R. F. Beers of the Geological Society of America and the Geotechnical Corporation; and to Dr. W. B. Nottingham of the Institute Department of Physics for their unstinted and practical assistance as well as moral encouragement.

# A Note on a Paper by Faust and Beck

WILLIAM M. STONE
*Iowa State College, Ames, Iowa*

(Received October 31, 1946)

An infinite sum transformation is defined and applied to a system of linear difference equations discussed by Faust and Beck in their paper on single tuned amplifiers. Some transforms of the more common functions are given and points of superiority of the transform method over the classical methods of solution of difference equations are emphasized.

---

IN their paper, "Oscillation Conditions in Single Tuned Amplifiers,"[1] Faust and Beck solved a simultaneous set of two first-order difference equations by classical methods. The purpose of this paper is to introduce the Generalized Laplace Transformation of Samuelson[2] and to illustrate its application to the same problem in circuit analysis. The advantages are that (1) a difference system is reduced to an algebraic system, (2) there are no extraneous constants introduced, and (3) no assumptions need be made as to the nature of the solution.

The generalized Laplace transformation is defined as

$$L\{F(k)\}_E = f(s) = \sum_0 F(n)s^{-n-1}, \tag{1}$$

where $F(k)$ must be of exponential order, $F(k) < M\alpha^k$, $k > K$, and defined for all positive integral values of $k$. The series is then uniformly convergent for all values of $s > \alpha$. The subscript refers to the shifting operator, $E^m F(k) = F(k+m)$, and serves to distinguish the generalized Laplace transformation from the ordinary case. Some transforms may be found by elementary summation but a more powerful method is readily available. By definition of the $E$ operator

$$\begin{aligned} L\{E\cdot F(k)\}_E &= \sum_0^\infty F(n+1)s^{-n-1}, \\ &= s[F(0)s^{-1} + F(1)s^{-2} + \cdots] - F(0), \\ &= sf(s) - F(0), \end{aligned} \tag{2}$$

which may be generalized at once by mathematical induction to the form

$$L\{E^m \cdot F(k)\}_E = s^m f(s) - \sum_{j=0}^{m-1} s^{m-1-j} F(j). \tag{3}$$

Following is a list of some of the more useful transform pairs:

| $F(k)$ | $f(s)$ |
|---|---|
| $1$ | $1/(s-1)$ |
| $a^k \cos wk$ | $\dfrac{s - a\cos w}{s^2 - 2as\cos w + a^2}$ |
| $a^k \sin wk$ | $\dfrac{a\sin w}{s^2 - 2as\cos w + a^2}$ |
| $a^k \cosh wk$ | $\dfrac{s - as\cosh w}{s^2 - 2as\cosh w + a^2}$ |
| $a^k \sinh wk$ | $\dfrac{a\sinh w}{s^2 - 2as\cosh w + a^2}$ |
| $a^k k^{(n)}$ | $a^n n!/(s-a)^{n+1}.$ |

Here $k^{(n)}$ means the Newton factorial polynomial, $k(k-1)(k-2)\cdots(k-n+1)$.

Using capital letters for the object function, small letters for the transforms, Faust and Beck's equations for voltage and current may be written as

$$E(k+1) - E(k) - Z_1 I(k) = 0, \tag{4}$$

$$E(k+1) + g_m Z_2 E(k) - Z_2 I(k+1) + Z_2 I(k) = 0. \tag{5}$$

By means of (2) these may be transformed at once to an algebraic system:

$$(s-1)e(s) - Z_1 i(s) = E(0), \tag{6}$$

$$(s + g_m Z_2)e(s) - Z_2(s-1)i(s) = E(0) - I(0)Z_2. \tag{7}$$

---

[1] Faust and Beck, J. App. Phys. 17, 749 (1946).

[2] Samuelson, Bull. Am. Math. Soc., Abstract 52-3-86 (1946).

The solution of this system is found to be

$$e(s)=\frac{E(0)\left[s-\left(1+\frac{Z_1}{2Z_2}\right)\right]+I(0)Z_1-E(0)\frac{Z_1}{2Z_2}}{s^2-2s\left(1+\frac{Z_1}{2Z_2}\right)+1-g_mZ_1},\tag{8}$$

$$i(s)=\frac{I(0)\left[s-\left(1-\frac{E(0)}{I(0)Z_2}\right)\right]+E(0)g_m}{s^2-2s\left(1+\frac{Z_1}{2Z_2}\right)+1-g_mZ_1}.\tag{9}$$

Substitute

$$a=(1-g_mZ_1)^{\frac{1}{2}},\quad \cosh w=\frac{1+\frac{Z_1}{2Z_2}}{a},\quad \sinh w=\frac{\left[\frac{Z_1}{Z_2}+\left(\frac{Z_1}{2Z_2}\right)^2+g_mZ_1\right]^{\frac{1}{2}}}{a},$$

$$=A/a,\qquad\qquad =B/a,$$

so that

$$e(s)=\frac{E(0)(s-A)+I(0)Z_1-E(0)\frac{Z_1}{2Z_2}}{s^2-2As+a^2},\tag{10}$$

$$i(s)=\frac{I(0)\left[s-\left(1-\frac{E(0)}{I(0)Z_2}\right)\right]+E(0)g_m}{s^2-2As+a^2}.\tag{11}$$

Making the denominators conform with the list of transforms the object functions may be written as

$$E(k)=a^{k/2}\left[E(0)\cosh wk+\frac{I(0)Z_1-E(0)\frac{Z_1}{2Z_2}}{B}\sinh wk\right],\tag{12}$$

$$I(k)=a^{k/2}\left[I(0)\cosh wk+\frac{I(0)\frac{Z_1}{2Z_2}+\frac{E(0)}{Z_2}(1+g_mZ_2)}{B}\sinh wk\right].\tag{13}$$

The boundary conditions are

$$I(n)=0,\tag{14}$$

$$E(0)=E_{in}+Z_gI(0),\tag{15}$$

where $I(n)$ is the current in the last mesh, $E(0)$ is the voltage in the input stage, $E_{in}$ is the voltage of the input generator whose internal impedance is $Z_g$. Substitution of both conditions in (13) and algebraic reduction leads to

$$E(0)=E_{in}\frac{B\cosh wn+\dfrac{Z_1}{2Z_2}\sinh wn}{B\cosh wn+\left[\dfrac{Z_1}{2Z_2}+Z_0\left(\dfrac{1}{Z_2}+g_m\right)\right]\sinh wn}, \tag{16}$$

$$I(0)=-E_{in}\frac{\left(\dfrac{1}{Z_2}+g_m\right)\sinh wn}{B\cosh wn+\left[\dfrac{Z_1}{2Z_2}+Z_0\left(\dfrac{1}{Z_2}+g_m\right)\right]\sinh wn}. \tag{17}$$

In practice the gain off all $n$ stages is of most importance. If $Z_0$ is set equal to $Z_2$ as is customary the gain of the $n$ stages may be reduced to

$$\frac{E(n)}{E_{in}}=\frac{B(1-g_mZ_1)^{n/2}}{B\cosh wn+(A+g_mZ_2)\sinh wn}. \tag{18}$$

The well-known expansion formulas

$$\cosh wn=\sum_k\binom{n}{2k}(\cosh w)^{n-2k}(\sinh w)^{2k}, \tag{19}$$

$$\sinh wn=\sum_k\binom{n}{2k+1}(\cosh w)^{n-2k-1}(\sinh w)^{2k+1}, \tag{20}$$

where $\binom{n}{2k}$ represents the binomial coefficients as $k$ takes on the values 0, 1, 2, ..., lead to the form of (18)

$$\frac{E(n)}{E_{in}}=\frac{(1-g_mZ_1)^n}{\sum_k\binom{n+1}{2k+1}A^{n-2k}B^{2k}+g_mZ_2\sum_k\binom{n}{2k+1}A^{n-2k-1}B^{2k}}, \tag{21}$$

which is equivalent to the general formula for the gain given by Faust and Beck.

# Internal Friction of Zinc Single Crystals*

IRVIN H. SWIFT†** AND JOHN E. RICHARDSON***
*Department of Physics, The State University of Iowa, Iowa City, Iowa*

(Received December 24, 1946)

**Measurements were made of the internal friction of single crystal rods when vibrating longitudinally at 45 kilocycles per second. The decrement, or ratio of energy dissipated per cycle to twice the total vibrational energy, was used as a measure of the internal friction. At strain amplitudes below about $10^{-5}$ or $10^{-6}$ the decrement was independent of strain amplitude. The value of the constant decrement at low amplitudes, and the changes which occurred as the amplitude was increased, were both greatly dependent on the history of the specimen. The constant decrement for different crystals ranged from $7\times 10^{-6}$ to values higher than could be measured (about $200\times 10^{-6}$). For some of the initially high-decrement crystals, values decreased with time (room temperature anneal). In general, for annealed crystals the decrement, measured as a function of increasing strain amplitude, remained almost constant up to and beyond the static elastic limit and then rose rapidly. It returned to the initial value when the strain amplitude was again reduced. On a second run, however, the rise in decrement occurred at much lower amplitude than on the initial one. It is concluded that the decrement is the result of two separate processes. The changing decrement at higher strain amplitudes is qualitatively explained with the use of dislocation theory.**

## INTRODUCTION

THE internal friction of single crystals of exceedingly pure zinc has been measured by Read,[1] and of less pure zinc by Read and Tyndall.[2] The decrement, which is the ratio of the energy dissipated per cycle to twice the total vibrational energy, was taken as a measure of the internal friction. Read found the decrement to be dependent on the past history of the specimen, strain or stress amplitude at which it was measured, time, etc. For the purer zinc decrements of the order of $10^{-5}$ were found at the lowest strain amplitudes but these, for most of the crystals, increased many fold with increasing amplitude. The present work consists of further observations of the behavior of the decrement of longitudinally oscillating zinc crystals under various circumstances and with partially controlled history.

## EXPERIMENTAL METHOD

The zinc[3] used in this investigation was slightly less pure than that used by Read.[1] The crystals were grown by the Bridgman method in which a small casting is lowered through a furnace. The majority of the crystals were grown with the casting imbedded in dental investment compound and the rest were grown in the glass tube in which the original casting was made. The dental investment compound was easily softened by water after the crystals were grown and could then be removed with little danger of straining the crystals. The crystals had a cross-sectional area of 9 sq. mm and had fairly smooth surfaces.

The decrements were measured by essentially the same method as that used by Read[4] in which a composite oscillator is constructed by cementing to a quartz crystal a zinc crystal of the same cross-sectional area and cut with a fine jeweler's saw to a length such that the composite oscillator resonates at a frequency within less than 4 percent of the resonant frequency of the quartz crystal. The resonant frequency of the quartz crystal used in the present work was 45 kilocycles when vibrating longitudinally in the fundamental mode. The composite oscillator was suspended by silk fibers placed at displacement nodes of vibration between, but not touching, parallel plane electrodes. The assembly was then placed in a jar and the jar evacuated to a pressure of less than

---

* The main contents of two dissertations submitted in partial fulfillment of the requirements for degrees of Doctor of Philosophy in the Department of Physics in the Graduate College of the State University of Iowa.

** April, 1943.

*** June, 1946.

† Now at the Naval Ordnance Test Station, Inyokern, California.

[1] T. A. Read, Phys. Rev. 58, 371 (1940).

[2] T. A. Read and E. P. T. Tyndall, J. App. Phys. 17, 713 (1946).

[3] The zinc used was "Bunker Hill," a commercial zinc with not more than 0.01 percent impurity. Read and Tyndall (reference 2) report data from three crystals grown from this same zinc and two with additional impurity.

[4] The writers are indebted to Dr. Read for information about his bridge circuit, etc., in addition to that in reference 1.

0.1 mm of mercury. The quartz crystal was etched with hydroflouric acid before using and had a decrement of $2.7\times 10^{-6}$.

The electrical apparatus consisted of an alternating-current bridge containing the composite oscillator as one arm and driven by a stable 35–100 kc oscillator. The detector used with the bridge consisted of an r-f amplifier, pentagrid converter, and associated oscillator which changed the high frequency signal from the bridge to an audible signal, and several stages of audio amplification. Its sensitivity with headphones was greater than that needed in any of the measurements. Either a vacuum thermocouple or a vacuum tube voltmeter were used to measure the voltage $\epsilon$ applied to the composite oscillator. The maximum strain amplitude, $U_0$, in the zinc specimen was then computed from the following formula:[5]

$$U_0=\frac{2\epsilon}{RV_0}(10K)^{\frac{1}{2}}\times 10^3, \tag{1}$$

where $R$ is the electrical resistance of the composite oscillator at resonance, $V_0$ is the velocity of sound in the zinc specimen, and $K$ is the constant relating the inductance $L$ of the composite oscillator and the masses $M_1$ and $M_2$ of the cylinders in the equation:

$$L=K(M_1+M_2). \tag{2}$$

The subscripts 1 and 2 in Eqs. (1) and (2) refer to the quartz and specimen, respectively.

The decrements of the zinc crystals were obtained from a knowledge of the inductance in the equivalent circuit, the resistance measured by the bridge, the resonant frequency of the composite oscillator, and the masses of the quartz and specimen. The constant $K$ was evaluated with Eq. (11) in a paper by Cooke[6] and Eq. (2), from the observed variation of the reactance of the quartz crystal alone with frequency in the immediate neighborhood of the resonance frequency. It was determined to an accuracy of 1.5 percent. The accurate measurement of changes in frequency required in the plotting of the resonance curve was obtained as follows. The crystal controlled multivibrators of a piezoelectric frequency standard were adjusted in frequency until a harmonic of one of them was within about 50 cycles of the resonance frequency of the quartz crystal. The difference frequency obtained from this reference frequency and the frequency of the oscillator driving the bridge was amplified and used to drive a Cenco counter. The number of counts per second was then equal to the deviation of the unknown frequency from the crystal controlled reference frequency.

## EXPERIMENTAL RESULTS

Before describing the quantitative results, some qualitative observations may be of interest. It was noted that some of the crystals could be made to give a high pitched "ringing" sound when set into flexural oscillations by suspending them from the mid-point and tapping lightly. Later work showed that crystals for which this ringing was audible for 2 or 3 seconds had decrements of the order of $10^{-4}$ or less, whereas higher decrement crystals instead of ringing yielded a metallic or even wooden sound. A rough idea of the behavior of the decrement of a crystal could be obtained by subjecting it to this simple test. For example, eight crystals were tested just after they had been sawed to the desired length and then again after an anneal at 250°C for three hours. No noticeable difference was obtained at this time, but after storage at room temperature for several days a noticeable drop in the decrement, as judged by the increased time of ringing, was observed for some of the crystals while the rest showed little change. The crystals of the former class seemed to be persistent in their behavior. For example, crystal No. 1 (orientation 15°) would always return to a low decrement after two or three days' rest at room temperature without handling, regardless of its previous history. At one time when it was known to have a low decrement, it was accidently dropped onto the floor and bent considerably. It was then straightened by hand and left idle for several days after which it appeared to ring as well as ever. The data in Fig. 1 (curve $A$) and Fig. 8 (curve $A$) were taken on this crystal after it had rested nine days before it was mounted in the

---

[5] This formula is an obvious modification of formula (1), reference 2; for stress amplitude. It may be obtained by equating the expression for total energy in electrical terms to that in mechanical terms, and solving for $U_0$.

[6] W. T. Cooke, Phys. Rev. 50, 1158 (1936). Cooke's $K$ is eight times the $K$ used in Eq. (2).

apparatus and 7½ hours afterward. Other crystals behaved similarily although none of these were subjected to such drastic treatment. Crystals which did not develop a ring in a few days after growth did not do so after months of lying idle. Several crystals of this type were also annealed at 250°C and still did not develop a ring. These qualitative tests moreover showed very clearly how little handling was needed to increase the decrement. If a crystal which had a good ring was placed on a table and one end was elevated about a fourth of an inch and then dropped, the crystal thereafter would give only a metallic "clink" when tapped. In fact, while the crystals were being subjected to the ringing test, the gentle taps needed to set up the oscillations a few times would noticeably reduce the time of ringing. One or two days of rest at room temperature after the above treatment was however sufficient to restore the ringing properties of the crystals.

Nothing was found in the method of growth, amount of handling, etc., which gave any clue as to why some crystals had a low decrement while others had decrements which were too high to measure with the apparatus and which would not decrease with either a high or room temperature anneal. There was no apparent correlation of high and low decrement crystals with the degree of smoothness of their surfaces.

The writers observed, as did Read,[1] that the decrement measured at low amplitudes immediately after mounting in the apparatus was always much higher than after the crystal had been allowed to rest in the apparatus for some time undisturbed except for infrequent low amplitude measurements.[7] Two examples of this behavior are given in Fig. 1. Curves *A* and *B* are for two different crystals. The decrease in decrement with time after mounting was caused by, at least primarily, the self-annealing of the crystals at room temperature. Part of the decrease, particularly the initial rapid decrease, may have been caused by the drying of the cement between the specimen and the quartz. Calculations based on estimates of the thickness of the cement layer, etc., indicate, however, that the decrement of the cement would have to be as large as 100 to 1000 to cause the high decrements obtained. This seems improbable in view of the fact that the cement had dried for twenty or thirty minutes before the crystals were mounted. Also, Fig. 2 in Read's paper[1] shows a similar decrease of the decrement with time after mounting and his crystals were fastened to the quartz with beeswax and rosin which would show no drying effect.

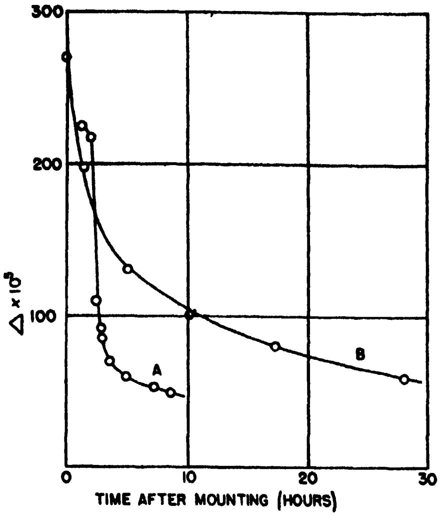


FIG. 1. Decrease of decrement with time after mounting, showing room temperature annealing. Curve *A* was taken on crystal No. 1.

Observations were made of the decrement as a function of maximum strain amplitude and time for thirteen crystals. It is not possible to present all of these data in full detail, but certain rather typical data are summarized and presented graphically. In Fig. 2 are shown curves of the decrement against strain amplitude for three crystals which had been allowed to rest in the apparatus for the following lengths of time: No. 2, two days; No. 3, three days; No. 4, thirty hours. In all cases the decrement had decreased to a nearly constant value before the data were taken. Crystal No. 4 (curve *C*) is particularly interesting as its decrement was practically constant to a value of maximum strain amplitude over five times that at the static elastic limit[2] (about $80 \times 10^{-7}$) for this crystal. Subsequent to the data in Fig. 2, crystal No. 4 was stored at room temperature for over six months. It was then remounted in the apparatus and the data

---

[7] The crystals of reference 2 made of the same zinc as the writers' did not show this decrease of decrement with time. These crystals however were of larger cross section than the writers' and the natural conclusion seems to be that they could be mounted without producing any detrimental strains.

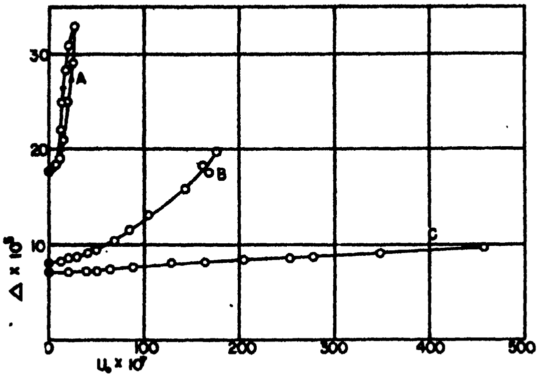


FIG. 2. Typical decrement *versus* maximum strain amplitude curves. Curve *A* was for crystal No. 2 $(\theta = 71^\circ)$; curve *B* was for crystal No. 3 $(\theta = 71^\circ)$; and curve *C* was for crystal No. 4 $(\theta = 70^\circ)$.

shown in Fig. 3 were taken. The decrement had been allowed to reach a fairly constant value after mounting before the data were taken. The time given with each curve is the elapsed time after the taking of the first curve. The decrement was measured for increasing and then decreasing strain amplitude, as shown by the arrows. The decrement at the lowest strain amplitude decreased from $95\times10^{-5}$ to $70\times10^{-5}$ in the course of 143 hours, but did not come down to the previously measured value of $7.0\times10^{-5}$ (Fig. 2). It may be seen from the lower branch of each of the curves shown in Fig. 3 that the strain amplitude required to increase the decrement by some amount, say 50 percent, increases considerably with elapsed time. In curve *J* the maximum strain amplitude was at the static elastic limit $(80\times10^{-7})$ before the decrement started to increase very rapidly. Where two curves were taken with little intervening time, such as *B* and *C*, *G* and *H*, and *I* and *J*, the ascending branch of the second curve of a pair was nearly identical to the descending branch of the previous curve. After these data were taken, the crystal was removed from the apparatus and allowed to rest for 20 days. It was then remounted in the apparatus and left idle for seven days after which the data shown in curve *B*, Fig. 4, were obtained. The decrement in this curve shows only a gradual rise to a strain amplitude nearly three times that at the elastic limit. The decrement at low strain amplitude is still however about nine times as large as the first previously measured value (curve *C*, Fig. 2).

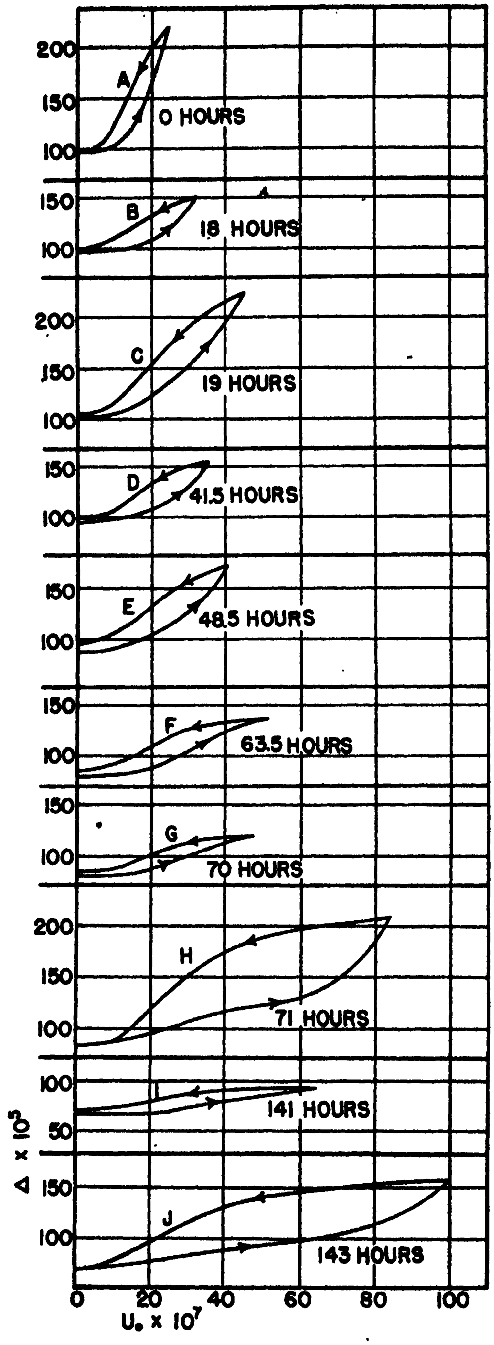


FIG. 3. Decrement *versus* maximum strain amplitude $U_0$ for crystal No. 4 $(\theta = 70^\circ)$. The time given by each curve is the elapsed time after the first curve was taken. These curves show the effect of self-annealing at room temperature.

The curves shown in Fig. 5 demonstrate very well the phenomenon mentioned in connection with Fig. 3. These curves were taken on crystal

No. 6 ($\theta = 56°$) which was known to be strained prior to mounting. Data were first taken to a strain amplitude of $50 \times 10^{-7}$ (curve 1) and the strain amplitude was then reduced to its lowest value. Curve 2 was then run to a strain amplitude of $80 \times 10^{-7}$, with curves 3 and 4 following in a similar fashion. From these curves it appears that previous runs affect subsequent runs only at strain amplitudes below the highest maximum strain amplitude obtained on the previous run. Thus, had the crystal been taken continuously from the minimum to the maximum strain amplitude ($94 \times 10^{-7}$), it would probably have followed the lowest curve, which is a smooth continuation of curve 1. Also, while descending curves are not shown, they coincide quite well with the next ascending curve, as previously pointed out in connection with Fig. 3.

The curves shown in Fig. 6 taken on crystal No. 5 (orientation 80°) illustrate some effects of room temperature annealing. The decrement of this crystal remained constant until a critical value of strain amplitude was obtained and then started to rise. As successive runs were taken, this critical value of strain amplitude increased. The data for curve *A* were taken soon after the crystal was mounted. Data for curve *B* were taken 53 hours after *A* was taken; and curve *C*, 98 hours after *B*. Other sets of data were taken on the crystal during these two periods and therefore they should not be considered strictly as rest periods. Finally the crystal refused to anneal any further at room temperature and data could be very nearly repeated on successive runs taken several days apart. These later curves were substantially identical to curve *C*, Fig. 6. At this point the crystal was annealed for several days at 260°C, mounted in the apparatus, and left idle for a week. The decrement then remained nearly constant for a range of stress amplitude 2½ times the value at the static elastic limit, as shown by curve *D*, Fig. 4. It will be noticed that after the high temperature anneal and remounting, the decrement at low strain amplitudes was higher than before the anneal (Fig. 6).

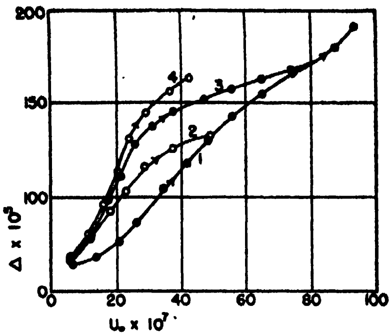


FIG. 5. Decrement *versus* maximum strain amplitude $U_0$ for crystal No. 6 ($\theta = 56°$), which was known to be strained before mounting in the apparatus. The curves are shown for increasing strain amplitude only and were taken in the order designated. They show that the decrement was influenced by previous runs only at strain amplitudes lower than the highest strain amplitude obtained on previous runs.

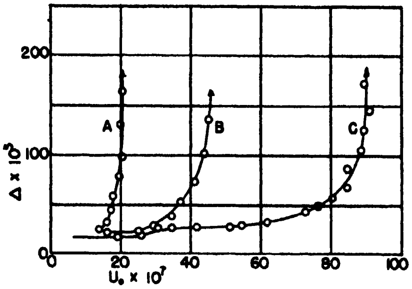


FIG. 6. Decrement *versus* maximum strain amplitude $U_0$ for crystal No. 5 ($\theta = 80°$). Curve *A* was taken soon after mounting, curve *B* 53 hours after *A*, and curve *C* 98 hours after *B*. Succeeding curves were very little different from curve *C*. As the crystal annealed at room temperature, the point of rise of the decrement occurred at higher strain amplitudes.

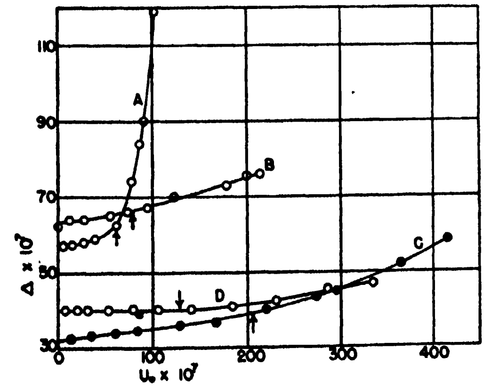


FIG. 4. Decrement *versus* maximum strain amplitude $U_0$ for crystals No. 9, 4, 10, and 5 corresponding to curves *A*, *B*, *C*, and *D*, respectively. All of these crystals had rested for a long period of time in the apparatus after mounting and before measurements were taken. The arrows on each curve designate the strain corresponding to the minimum static stress which will produce glide in the basal plane.

The two curves given in Fig. 7 indicate the diverse results that can be obtained when the conditions are very nearly the same. Curve *A* is

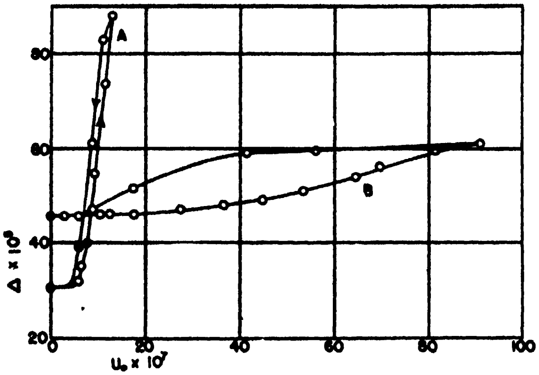


FIG. 7. Decrement *versus* maximum strain amplitude $U_0$. Curve $A$ was for crystal No. 7a and curve $B$ was for crystal No. 7b. Each of these two crystals were sections cut from one long 60° crystal.

for crystal No. 7a and curve $B$ is for No. 7b. These two crystals were sections cut from the same original 60° crystal and annealed for two hours at 250°C. Crystal No. 7a was mounted four days after the anneal and curve $A$ was taken after it had rested in the apparatus for one day. Crystal No. 7b was mounted eleven days after the anneal and curve $B$ was taken after it had rested in the apparatus for two days. Curve $A$ rises sharply at a low value of strain amplitude, while curve $B$ shows a constant decrement as the strain amplitude is increased to a value of six or seven times that at which curve $A$ starts to rise. The only known difference in the histories of the two crystals was in the time they were allowed to rest before and after mounting in the apparatus. Crystal No. 7b apparently had sufficient time to anneal at room temperature while No. 7a was still in a strained state. One other curve was taken on crystal No. 7b one day previous to curve $B$ (one day after mounting), and it was nearly identical to curve $B$ indicating that this crystal was in an annealed state (see discussion of Fig. 6).

Curve $A$ of Fig. 8 was taken on the crystal (crystal No. 1, orientation 15°) which was dropped, bent, and subsequently straightened, as mentioned earlier. After the above treatment it was left idle for five days and mounted in the apparatus. At this time its resonant frequency was found to be too low and it was necessary to cut it to a shorter length. Three days after the sawing operation it was remounted in the apparatus and left idle for about eight hours except for occasional measurements of the decrement at very low strain amplitudes (curve $A$, Fig. 1). The data for curve $A$, Fig. 8 were then taken. Curve $B$ represents data on crystal No. 8 (orientation 73°). It had rested for several days before mounting, and 34 hours after, before the data were taken. These two crystals had about the same resolved shear stress in the basal plane for a given tensile stress. The two curves are similar in form, indicating that crystal No. 1 had recovered fairly well from the severe strain given it.

The curves $A$, $B$, $C$, and $D$ in Fig. 4 are for crystals which had been allowed to rest for long periods of time in the apparatus after they were mounted. The histories of crystals No. 4 and No. 5 have been given earlier in this paper. Crystal No. 9 (orientation 50°) and crystal No. 10 (orientation 20°) were left idle in the apparatus for about the same periods of time as crystals No. 4 and No. 5. The decrement of three of the crystals rose slowly with strain amplitude for values of strain amplitude considerably greater than the value of the static elastic limit, while the decrement of the other crystal began to rise sharply at about the static elastic limit (curve $A$). This crystal (orientation 50°) probably was still in a strained state as crystals with orientations near 45° are much more easily strained than those with orientation angles near 0° or 90°.

When the voltage across the composite oscillator was suddenly changed from one constant value to another, the decrement took an appreciable time to change from the previous value

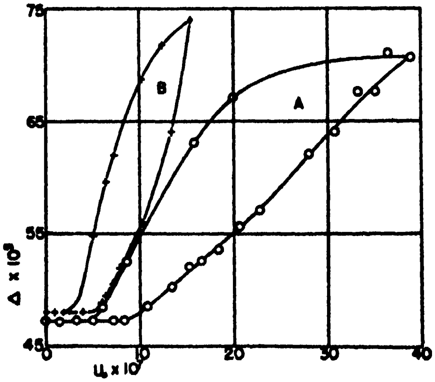


FIG. 8. Decrement *versus* maximum strain amplitude $U_0$. Curve $A$ was for crystal No. 1 and curve $B$ was for crystal No. 8.

to the final value at the new stress amplitude. The curves in Fig. 9 illustrate this effect. For curves 1 and 2 the voltage change was an increase while for the others the voltage was decreased. Usually when the voltage across the crystal was changed, a time of less than 20 seconds was required for the decrement to attain, say, 90 percent of its total change. In two of the curves of Fig. 9, however, the decrement required about a minute to attain 90 percent of its change. No attempt was made to fit these data with a mathematical function because the data were too inaccurate, particularly in the region where the decrement was changing rapidly with time. In addition there were other complicating factors. The strain amplitude is proportional to the current through the crystal and hence as the decrement (proportional to resistance, which determines the current) changed with time after the voltage across the crystal was changed to the new value, the strain amplitude was not constant during the time indicated in Fig. 9. Also, because of the low decrements involved, the time required for the strain amplitude to reach a steady state upon a change of the applied voltage from one value to another was large enough to make the effective zero of time uncertain.

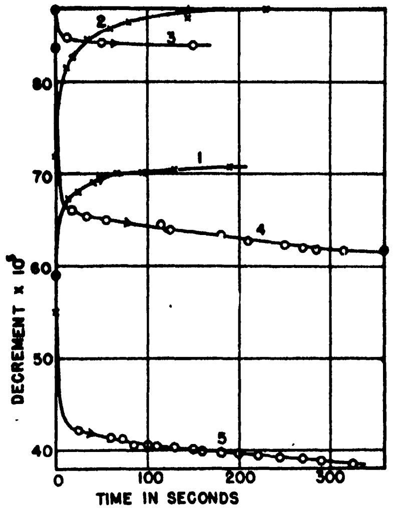


FIG. 9. Curves showing for crystal No. 7a the change in decrement with time following a sudden change of driving voltage across the composite piezoelectric oscillator. The curves were taken in the order designated by the numbers on each curve.

### DISCUSSION OF DATA

A number of points should be mentioned in regard to the measurements described above. These are:

(1) The decrement measured at low amplitude decreased with time after mounting in the apparatus, approaching a constant value (Fig. 1). The final value reached was not, however, unique, in that a crystal after a second mounting would show a decline in decrement to a constant value but not necessarily to the same value found in the first experiment.

(2) The decrement was independent of strain amplitude for all maximum strain amplitudes below a critical value of the order of $10^{-5}$ to $10^{-6}$ (the bridge detector was sufficiently sensitive to balance the bridge at strain amplitudes much lower than $10^{-7}$, probably $10^{-9}$). The value of the low amplitude decrement was, for a given mounting in the apparatus, affected only by time (as shown in Fig. 1).

(3) The value of the low amplitude decrement was higher for all crystals measured than for Read's[1] spectroscopically pure zinc crystals. The lowest decrements found were about the same as for those of crystals made of the same zinc, as reported by Read and Tyndall.[2] It does not follow, however, that the decrement is a function only of the amount of impurities present. Other factors are much more important as is shown by the data on crystal No. 4. In Fig. 2 it had a decrement of $7\times10^{-5}$ whereas after it had lain idle for six months and was remounted in the apparatus, the decrement at low amplitudes was $70\times10^{-5}$ (Fig. 3).

(4) The value of strain amplitude at which the decrement started to increase became greater with elapsed time following mounting in the apparatus, i.e., increased as the value of the low amplitude decrement decreased. This effect is illustrated in Figs. 3 and 6.

(5) If a crystal had been oscillated at a sufficient amplitude momentarily to increase the decrement, then on measurements made soon after, the decrement was larger at all strain amplitudes between the value at which the decrement started to rise and the highest value obtained on the previous run. The amount that a run influenced the decrement obtained on a succeeding run decreased with elapsed time between the runs. The effect of time between runs is

illustrated particularly well by Fig. 5 but is also evident from the series of curves in Fig. 3.

(6) The decrement lagged in time behind the oscillation amplitude when the amplitude was changed suddenly. It is apparent from these curves that a finite time was required for the decrement to reach a stable value after the oscillation amplitude was changed.

The shape of the curves of decrement *versus* strain amplitude suggests that the observed decrement was composed of a constant decrement plus an amount which varied with strain amplitude. The portion of the decrement (to be designated by decrement $A$) which was constant with strain amplitude was measured alone at small strain amplitudes. Decrement $B$ is that portion which was zero or negligible at zero strain amplitude and increased with increase of strain amplitude. It is possible on this basis to give a qualitative explanation of the behavior of the decrement in the experiments described above. The dislocation theory is used together with several assumptions to be listed later.

The dislocation theory[8] is based upon a certain type of local imperfection in the crystal lattice called a line dislocation. The local imperfection is concentrated about a line in a slip plane and is composed of such an arrangement of atoms that movement of the line imperfection through the crystal along a slip plane gives the same effect as if the material above the slip plane had been shifted a distance of one atomic spacing with respect to that below. The dislocations can be either positive or negative, one being complementary to the other in the sense that it is the mirror image of the other. A dislocation experiences a force in a shear stress field which reverses in sign with change in sign of the dislocation. Hence since a dislocation sets up a shear stress field of its own, dislocations of unlike sign are attracted to each other and those of like sign are repelled. Also a dislocation is attracted to the crystal surface by the force which would be given by an image dislocation of opposite sign outside of the surface. Dislocations can be formed in pairs in the interior of a crystal or singly at the surface as the result of shear stress.

In addition to the properties of dislocations summarized briefly above, the following assumptions will be used in the interpretation to be given for the results:

(a) The dislocations present as a result of previous low frequency, or "static" strains tend to form a stable lattice-like array within the crystal specimen.[8]

(b) The low amplitude decrement ($A$) is a function only of the total number of dislocations present at a given time in the specimen.

(c) The number of dislocations, if any, that are formed in the specimen as a result of the oscillating stress is a negligible fraction of the total number present. Presumably the short time during which stress was applied did not allow much opportunity for dislocations to form either pairwise in the interior, or at the surface.

(d) The value of decrement $B$ is a function of the number of dislocations present which are "free" to dissipate energy. Whenever a dislocation is held within a dislocation lattice it is assumed to mean that the dislocation is prevented, or at least restricted, from participation in the energy dissipation process.

A crystal which had been recently mounted in the measuring apparatus would be expected to be in a strained state as a result of handling during mounting. Hence there would be a large number of dislocations of both signs distributed throughout the crystal. With the aid of thermal fluctuations, the forces on and between dislocations would cause them to migrate. Some dislocations, particularly those near the surface, would diffuse out of the crystal, others would disappear by cancellation of positive and negative dislocations (if the pair was originally in the same slip plane), and the remaining ones would migrate to positions such that the crystal had a minimum of potential energy, i.e. they would arrange themselves in some sort of orderly pattern. After a long time the number of dislocations in the crystal would be essentially constant, all being held in the lattice. Under assumption (b), that decrement $A$ is a function only of the total number of dislocations present, the curves of Fig. 1 would follow. Furthermore, the final value reached by the decrement several days after

---

[8] A summary of the dislocation theory is given in the article by Frederick Seitz and T. A. Read, J. App. Phys. 12, 100 (1941). The forces between certain types of dislocations and other related matters are discussed in J. S. Koehler, Phys. Rev. 60, 397 (1941).

mounting would not be expected to be the same for a different mounting in the apparatus. Each mounting would entail a different kind and amount of the unavoidable strain, resulting in a different number and distribution of dislocations immediately thereafter. Hence the stable lattice reached would not be unique. This explains such data as that in Fig. 2 (crystal No. 4) where the low amplitude decrement was $7\times10^{-5}$ and later after the crystal was remounted the decrement was $70\times10^{-5}$, as shown in Fig. 3. The process by which energy is dissipated in an amount proportional to the total number of dislocations and independent of amplitude is difficult to determine without more detailed information. One possibility is that irreversible heat flow occurs as the result of oscillating stress in an elastically inhomogeneous medium. In this event, the dislocations would presumably provide the elastic inhomogeneity. Quantative calculations should show whether or not this possibility is reasonable.

Decrement $B$, a function of strain amplitude, could be interpreted according to the following picture. The dissipation represented by decrement $B$ would be assumed to be the result of movement of the dislocations. This movement would result in a plastic change in length of the specimen, with consequent loss of oscillation energy. The data could then be interpreted by assuming that the dislocations did not move appreciably when they were attached to the dislocation lattice and that they could be broken away from the lattice by the application of a sufficiently large stress. The amount of stress required would vary according to how tightly each dislocation was bound in the lattice. Hence as the oscillation amplitude was increased steadily from zero, more and more dislocations would be freed from the restraining influence of their neighbors and be made available to dissipate energy, with a consequent rise of the decrement. The decrement, during the subsequent decrease of amplitude would then be larger at a given amplitude than on the ascending curve because more dislocations would be free than before. Examples of this hysteresis-like effect are given in Figs. 2, 3, 5, 7, and 8. This picture, furthermore, would explain the effect shown in Figs. 3 and 5 in which an ascending branch of a curve was nearly identical to the descending branch of a curve taken soon before it. It would be due to the fact that the same number of dislocations were free to dissipate energy on both curves. The dislocations that had been freed from the dislocation lattice by the oscillating stress would, upon removal of the stress, gradually take either their old positions, or new positions in the lattice. This process would, however, take a finite time as thermal energies would be expected to be an important factor. The data given in Fig. 3 show how elapsed time after a run erases the effect of a previous run. Also, the increase of the point of rise of the decrement shown in Fig. 6 would be caused by the fact that as the crystal had more time to anneal, the dislocations which were loosely bound in the lattice would disappear from the crystal, leaving only those which were tightly bound and required a large stress to make them available to dissipate energy.

The number of dislocations that are freed from the lattice would be expected to depend somewhat on the time that the crystal was oscillated at a given amplitude, i.e., some time would be required to establish an equilibrium state. The curves of Fig. 9 show that several seconds were required for a stable value of the decrement to be reached when the oscillation amplitude was changed from one constant value to another.

In conclusion, the writers wish to express their appreciation to Professor E. P. T. Tyndall for his guidance and help during the progress of this work.

# Letters to the Editor

## Measurement of Surface Tension

ROBERT WEIL
*South-West Essex Technical College and School of Art, Walthamstow, England*
January 27, 1947

SUGDEN'S[1] method was employed for measuring the surface tension of distilled water at a temperature of 25°C in order to determine how the value would be affected by imperfections in the circularity of the capillaries. The procedure was as follows: six capillaries were drawn out from pieces of glass tubing so that the diameter at the tip was between .04 and .06 cm. While still soft the tips were squeezed to different degrees; thus the cross sections assumed elliptical shapes. It was argued that the ratio $R$ of the minimum to the maximum diameter was a measure of the deviation from the true circularity. This ratio, as determined by means of a cathetometer, varied from .537 to .914. The diameter of the large tube was .2814 cm.

The maximum bubble pressures were determined for each specimen. A high rate of bubbling was used at first with smaller rates following it. The mean of three readings for each rate was taken: the divergences were negligible. No automatic counting device being available, it was not possible to exceed a rate of about 140 to 150 bubbles per minute. Graphs were drawn plotting pressure against rate and extrapolated to infinitely long periods of bubble formation. This extrapolation seemed justified since, after some practice, it became possible to reduce the rates to as small values as two bubbles per minute. A mean value of the measured radii was used in order to calculate the surface tension.

Two interesting results have been found: for $.9<R<1$ values for the surface tension agree very well with data as obtained by other workers with capillaries whose cross section approximates to a circle to a greater degree than those used in this investigation.

It would appear that Martin's[2] conclusions obtained from work on similar deviations in Jäger's original method do not apply here.

Secondly, when plotting the pressure $p$ against the rate of bubbling, it was found in all cases that the values of $p$ increased with the rate of bubbling up to about 60 bubbles per minute or more, and then definitely decreased for higher rates. It was found, as might be expected, that, on raising the rate to a greater value than could be measured visually, the pressure rose considerably. The cause for this kink in the curves is not quite obvious. It is a matter of some conjecture whether it can be accounted for by the fact that at small rates actual maxima and minima are recorded in the positions of the liquid in the gauge, while at great frequencies it is the mean positions which are observed. In some unpublished work L. B. Wood points out the necessity of extrapolating to infinitely long periods of bubble formation to obtain reliable and consistent results. It would seem that there are occasions when it is not sufficient to maintain the rate of bubbling at less than two per second, as is suggested by Adam.[3]

It was not possible to extend the measurements to capillaries with different diameters and different wall thicknesses: it is certainly desirable that this should be done.

---

[1] Sugden, J. Chem. Soc. p. 858 (1922); p. 27 (1924).

[2] Martin, Sitz. Akad. Wiss. Wien 123, 2491 (1914).

[3] Adam, *The Physics and Chemistry of Surfaces*, p. 375.

## Letter to the Editor

FRANCIS J. MURRAY
*The Institute for Advanced Study, School of Mathematics, Princeton, New Jersey*
January 8, 1947

MAY I respectfully call your attention to the fact that the suggestion of Mr. Eaglesfield on linear equation solvers, on page 1125 of the current issue of the *Journal of Applied Physics* (December, 1946) duplicates a paper of mine presented to the American Mathematical Society, September 1945, and abstracted in the November issue of the *Bulletin of the American Mathematical Society*, Vol. 51 (1945), page 883.

The content of the paper appears in full in my book *The Theory of Mathematical Machines*, a current publication of the Columbia University Press.

In this paper, I prove the convergence in the case in which $\Sigma\, d_i^2$ is used. In the case of a machine which uses $\max d_i$ as an indication, one can give examples where the machine will reach a minimum not zero.

I have constructed a number of models upon the $\Sigma\, d_i^2$ principle and a large one is now being constructed by the Watson Scientific Computing Laboratories at Columbia University.

# Here and There

### Personnel

The National Bureau of Standards recently announced the retirement of two longtime employees. Harvey L. Curtis, physicist at the Bureau since 1907, had deferred his retirement in 1945 to continue direction of interior ballistics research. Frederick J. Bates, sugar physicist who had also been chief of the Optics Division since 1941, has given more than 43 years of continuous service to the Bureau.

On December 30 Carroll L. Wilson resigned as vice president and director of National Research Corporation in order to assume his duties as general manager of the Atomic Energy Commission.

Ross Gunn has resigned as superintendent of the Physics Division, Naval Research Laboratory, to act as director of physical research for the U. S. Weather Bureau.

In January three appointments to the new Department of Engineering Sciences and Applied Physics at Harvard

University were announced. Leon N. Brillouin was appointed Gordon McKay Professor of Applied Mathematics, and James B. Fisk and Ronold W. P. King were named Gordon McKay Professors of Applied Physics.

Chester Peterson has been appointed physicist in the Resistance Measurements Section of the National Bureau of Standards.

The following men have new appointments to the Department of Physics at the University of Minnesota: Charles A. Critchfield, Frank Oppenheimer, Edward J. Lofgren, Joseph Weinberg, and Otto H. Schmitt.

### Division of High-Polymer Physics

The following are officers for 1947 of the Division of High-Polymer Physics of the American Physical Society: Chairman, Lawrence A. Wood; Vice Chairman, Hubert M. James; Secretary, W. James Lyons; Treasurer, Robert S. Spencer. Professor James and Dr. Spencer were also elected to the executive committee.

### General Electric Fellowships at Union College

For the third year General Electric Company is offering 50 six-week, all-expense fellowships to high school science teachers in seventeen northeastern states, for summer study at Union College in Schenectady, New York.

### Leonard Loeb Honored

*School and Society* reports that Leonard B. Loeb, professor of physics, University of California (Berkeley), has been chosen to deliver the 38th annual Kelvin lecture at the Institution of Electrical Engineers, London, England, on April 24. Professor Loeb will speak on "Electrical Discharge through Gases."

### Junior Chamber Puts Physicist in Top List

The U. S. Junior Chamber of Commerce published in January its selection of the nation's ten outstanding young men of the year. Included among them was Philip Morrison of Cornell University, atomic physicist, "for his outstanding work in the development of the atomic bomb and his struggle to educate the people to the danger and promise of atomic energy."

### Electronics Fellowships at MIT

A number of graduate and advanced research fellowships are being offered by the Massachusetts Institute of Technology, Cambridge, Massachusetts, for study and research in the field of electronics. They will be known as Industrial Fellowships in Electronics and are sponsored jointly by a group of industrial organizations concerned with the advancement of electronics and its applications. Applicants should communicate with the Director, Research Laboratory of Electronics. Under normal circumstances application must be made at least four months prior to the intended date of entrance.

### Work to Start Soon on Schenectady Atomic Laboratory

Construction of the Knolls Atomic Power Laboratory to be built near Schenectady, New York, by the Atomic Energy Commission will start this spring. The laboratory is being erected under the supervision of the General Electric Company, which will operate it for the government when completed. It is predicted the new building will be occupied by the middle of 1948.

### *Acta Crystallographica*

A special international organization, formed largely for this purpose, is planning the publication of a journal to perform the services formerly rendered by the *Zeitschrift für Kristallographie. Acta Crystallographica*, as it will be called, will be published either in England or America, and while French, German, and Russian will be permitted languages, the bulk of the journal will be in English. It is expected that there will be six issues per year totaling approximately 1000 pages. The price to individual subscribers is tentatively set at $10 a year.

The American representatives in the international organization are extremely anxious to obtain a good idea of the probable circulation of the journal in North America. Any reader of this notice who would, or who might, subscribe to *Acta Crystallographica* is urged to send a note or postcard to that effect addressed to Dr. Henry A. Barton, American Institute of Physics, 57 East 55 Street, New York 22, New York.

### Reorganization of American Institute of Physics

Reorganization of the American Institute of Physics and plans for a new semi-popular journal devoted to physics and its relation to society have been announced by the Institute. Effective immediately, the 7000 members of five societies (The American Physical Society, The Optical Society of America, The Acoustical Society of America, The American Association of Physics Teachers, and The Society of Rheology) will become members of the Institute. Heretofore the Institute has been an organization of these five Member Societies without any individual memberships. Provision is also made for associate membership in the Institute, open to others interested in physics. No change is made in the organization, functions, and activities of the five societies.

The new journal, for which initial financing is now being sought, will contain news about physicists, their meetings, their work, and other activities, and about similar features in related fields; information about the action of governmental and other non-physics agencies affecting physics; articles and letters presenting the views of physicists on problems of general importance; references to current literature in physics; book reviews, etc. Technical journals in the field of physics already being published by the Institute and its Member Societies will remain unchanged.

### Division of Electron and Ion Optics

The Division of Electron and Ion Optics will hold a meeting in connection with the Washington meeting of the American Physical Society on May 1–3, 1947. There will be a program of invited papers in the general field of "The Physics of Surface Phenomena." Contributed papers will be welcomed. For further information address J. R. Pierce (secretary-treasurer), Bell Telephone Laboratories, Inc., 463 West Street, New York 14, New York.

### The Bulletin of Mathematical Biophysics

The following is the table of contents for the March 1947 issue of *The Bulletin of Mathematical Biophysics:*

**Suggestions for a Mathematical Biophysics of Some Psychoses**—N. RASHEVSKY

**A Problem in Mathematical Biophysics of Interaction of Two or More Individuals Which May Be of Interest in Mathematical Sociology**—N. RASHEVSKY

**Mathematical Theory of Motivation Interactions of Two Individuals:** I.—ANATOL RAPOPORT

**The Mechanism of the Middle Ear: I. The Two Piston Problem**—MARTINUS H. M. ESSER

The University of Chicago Press, Chicago, Illinois
Volume 9 Number 1.

### Symposium at Ohio State

A Symposium on Molecular Structure and Spectroscopy will be held at the Mendenhall Laboratory of Physics at The Ohio State University from Monday, June 9, through Saturday, June 14. There will be discussions of the interpretation of molecular spectroscopic data as well as methods of obtaining such data. In addition, there will be sessions devoted to microwave and Raman spectroscopy. A dormitory will be available for those who wish to reside on the campus during the meeting. For further information, or for a copy of the program when it becomes available, write to Professor Harald H. Nielsen, Mendenhall Laboratory of Physics, The Ohio State University, Columbus 10, Ohio.

### International Nickel Cooperative Educational Program

A broadening of International Nickel Company's cooperation with universities and colleges in the United States and Canada in the field of engineering education through the distribution of technical literature has been announced by T. H. Wickenden, manager of the development and research division of The International Nickel Company, Inc.

The program will make available useful material for classroom instruction in training students in scientific fields. It has been offered to, and accepted by, a number of institutions in the United States and Canada which give accredited courses in mining, metallurgy, chemical engineering, and one or more other engineering courses. As rapidly as possible the program will be offered to all engineering schools.

The development and research division of the company will furnish to each institution an exhibit containing approximately 50 specimens of nickel-containing materials; a portable metals identification kit containing approximately 35 specimens of important metals and alloys for qualitative identification of metals and alloys; literature and other information concerning nickel and its alloys; and other data. Motion pictures of the company's mining, smelting, and refining operations will be made available.

# New Books

## Circuit Analysis by Laboratory Methods

BY CARL E. SKRODER AND M. STANLEY HELM. Pp. 288+xvi, 6×9 in. Prentice-Hall, Inc., New York, 1946. Price $5.35.

Some twenty years ago Professor Harry E. Clifford planned a system of laboratory reports which was used under his cognizance at M.I.T. and Harvard, and the form of which is still used there and in many other schools. The purpose of this system was to encourage the student to plan and think about the experiment rather than treat it as a mere manual exercise. The modus operandi was to require a preliminary report which involved the following sections: a. A statement of the problem, b. The theory involved, c. The procedure, d. The circuit diagrams, e. The results expected together with a sample calculation, and f. A proposed analysis of the results.

After having this preliminary report, together with a list of needed instruments, accepted by the instructor, the student was then free to perform the experiment and was expected to perform it with little additional guidance. The experiment was finished only when a final report, including the completion of Sections e and f, was accepted. This method was and still is in sharp contrast with the cook book method prevalent in some schools. This latter method is sometimes reduced to a point where the student is asked only to read instruments and to put the readings in appropriate squares on a data sheet provided by the instructor. There is no question as to which method does the better job of teaching.

In the book *Circuit Analysis by Laboratory Methods*, the authors attempt to follow the preliminary report system and at the same time reduce some of the drudgery involved in merely copying sections of theory from references. This they do admirably. However, by the title one is rather led to expect that the book will contain an entirely new and different set of experiments. This is not so, since many of the experiments are good old standbys, but the emphasis has very definitely been shifted from machines to circuits. Only three out of the thirty-five experiments are on rotating machine characteristics. Eight of the remainder are on series and parallel circuits including resonance, which acknowledges their importance both in engineering and as a teaching mechanism. Three are on polyphase circuits, and three on reactive power. The remainder are distributed among instruments, Kirchhoff's laws, power, resistance, etc.

Although the text explaining the philosophy of the experiments is good, it is expected that the course would be taken concurrently with a theory course in circuits. Since there should be some correlation between this and the laboratory, one would hope to find a classroom text which covers the same material in the same order. This might not be easy, although it might be accomplished by a certain amount of reorganization.

The chapter describing the report to be expected of the student is good, but it could have been rendered more vivid by including an example of what the authors consider to be a good report on one of the experiments included in the later chapters.

It seems that the authors have lost some crispness in trying to make the book so general that it can be used in all laboratories with all kinds of equipment. Somehow the sentence, "The inductance of the coil and the capacitance of the condenser should be such that resonance will be obtained at a frequency about in the middle of the frequency range designated by the instructor," does not seem nearly as authoritative as, "Choose the inductance and capacitance so that resonance will be obtained at 2000 cps."

ERIC A. WALKER
*The Pennsylvania State College*

### The Diffraction of X-Rays and Electrons by Free Molecules

BY M. H. PIRENNE. Pp. 160+xii, Figs. 82, $5\frac{1}{4}\times 8\frac{1}{4}$ in. Cambridge Series of Physical Chemistry. Cambridge University Press, Cambridge, England, 1946.

The thirteen chapters cover the following topics: Coherent scattering, incoherent scattering, fluorescence radiation, experimental measurements of the scattering of x-rays by atoms, diffraction of fast electrons by atoms, diffraction of x-rays by free molecules, intermolecular interferences in gases and liquids, thermal movements in molecules, the chemical bond, use of x-ray and electron diffraction for determination of structure, Fourier analysis of interference measurements of free molecules, measurement of x-ray scattering functions of gases, and a list of molecules studied by x-ray diffraction. An excellent bibliography contains 187 references.

The book is essentially a report of the work of the Debye school in the field of x-ray scattering by free molecules. The material is well presented and the book makes easy and interesting reading. Theoretical considerations and derivations are summarized, the results and conclusions are clearly stated, and the step-by-step details omitted. The first twelve chapters could well serve as a complete coverage for twelve semi-popular lectures. Most of the work presented in the book is more than ten years old, and much of it is familiar to workers in the field. However, even the experienced x-ray diffraction worker will find that the careful authoritative presentation clears up for him a few misunderstood points. The book should serve a useful purpose as a general report and reference for this phase of the x-ray diffraction field.

B. E. WARREN
*Massachusetts Institute of Technology*

## New Booklets

Tracerlab, Inc., 55 Oliver Street, Boston 10, Massachusetts, has published the first issue of *Tracerlog*, the purpose of which is "to publish technical information on laboratory, medical, and industrial uses of radioactivity, not to mention an occasional 'commercial' plug for Tracerlab's own products and services." It will be published monthly and mailed free of charge on request to physicists, chemists, physicians, and engineers interested in radioactivity and its applications.

Tube Department of Radio Corporation of America has announced the publication of two 16-page booklets. Form PG-101 is entitled "Power and Gas Tubes for Radio and for Industry" and Form 1275-C is named "Receiving Tubes for Television, FM, and Standard Broadcast." Either booklet may be obtained for 10 cents a copy from RCA tube distributors, or by sending 10 cents direct to Commercial Engineering, Tube Department, Radio Corporation of America, Harrison, New Jersey.

Miniature Precision Bearings, Inc., Keene, New Hampshire, offers a four-page bulletin on miniature ball bearings for precision instruments and mechanisms. It describes in detail five series: radial, super-light radial, pivot, angular contact, and thrust.

Acme Scientific Company, 1448 West Randolph Street, Chicago 7, Illinois, has issued a four-page folder describing in detail Acme's new mold polishing service. The company feels that the type of surface which it can provide on irregular contours may be useful to the readers of this journal.

Eberbach and Son Company, Ann Arbor, Michigan, publishes a 16-page house organ called *Announcer of Scientific Equipment*, of interest to research, control, and development laboratories.

Rex Rheostat Company, P. O. Box 232, Baldwin, Long Island, New York, announces its new *Catalogue No. 3*, 8 pages. The company manufactures slide-contact rheostats and resistors.

Leeds and Northrup Company, 4934 Stenton Avenue, Philadelphia 44, Pennsylvania, has published a new catalog, *ND46(1)*, which provides information about a pyrometer which supplements the Micromax line of instruments for those applications which require split-second response to temperature changes, or unusually rapid concentration on a single chart of data from many points. 16 pages.

Walker-Jimieson, Inc., 311 South Western Avenue, Chicago 12, Illinois, makers of radio and electronic supplies, have issued their 1947 *Reference Book and Buyer's Guide*. 106 pages.

Radio City Products Company, Inc., 127 West 26 Street, New York 1, New York, is announcing its new *Catalogue No. 129*, which describes a comprehensive line of radio, electrical, and electronic test instruments. 24 pages.

American Documentation Institute, 1719 N Street, N. W., Washington 6, D. C., has published its first *Catalogue* *of Auxiliary Publications in Microfilms and Photoprints.* To be distributed free so long as this first edition is available. 52 pages.

Dow Corning Corporation, Midland, Michigan, has announced the third edition of its *DC Silicone Catalogue.* 12 pages.

RCA Victor Division, Camden, New Jersey, has published two new booklets for the wider dissemination of technical information about electron tubes. Each contains 16 pages, and the titles are *Receiving Tubes for Television, FM, and Standard Broadcast*, and *Power and Gas Tubes for Radio and for Industry.*

Castle Films Division of United World Films, Inc., 30 Rockefeller Plaza, New York 20, New York, is announcing a new catalog listing many new films released during the past year for school and industrial training purposes.

The Ruberoid Company, 500 Fifth Avenue, New York 18, New York, has issued a 40-page illustrated booklet by Dr. Oliver Bowles entitled *Asbestos—The Silk of the Mineral Kingdom.* Dr. Bowles is chief of the Nonmetal Economics Division of the U. S. Bureau of Mines and is a prominent authority on the subject of asbestos.

Philips Research Laboratory has published a 20-page index listing all articles which have appeared in *Philips Technical Review* during the period January 1936 (first issue) to June 1942 (last issue prior to suspension of publication during the German occupation of Holland). It is available on request from Elsevier Book Company, Inc., 215 Fourth Avenue, New York 3, New York. The index, distributed free as a service to readers, alphabetically catalogs by subject all articles which appeared during that period.

## Calendar of Meetings

| | |
|---|---|
| **April** | |
| 28–30 | American Geophysical Union, Washington, D. C. |
| 28–30 | National Academy of Sciences, Washington, D. C. |
| **May** | |
| 1–3 | American Physical Society, Washington, D. C. |
| 8–10 | Acoustical Society of America, New York, New York |
| 17 | New England Radio Engineering Meeting, Cambridge, Massachusetts |
| 26–27 | Institute of Aeronautical Sciences, Detroit, Michigan |
| **June** | |
| 1–6 | Society of Automotive Engineers, French Lick, Indiana |
| 8–10 | American Society of Refrigerating Engineers, Los Angeles, California |
| 9–13 | American Institute of Electrical Engineers, Montreal, Quebec, Canada |
| 9–14 | Symposium on Molecular Structure and Microscopy, Ohio State University, Columbus, Ohio |
| 16–19 | American Society of Mechanical Engineers, Chicago, Illinois |
| 16–20 | American Society for Testing Materials, Atlantic City, New Jersey |
| 18–21 | Society for the Promotion of Engineering Education, Minneapolis, Minnesota |
| 23–25 | American Society for X-Ray and Electron Diffraction, Montreal, Quebec, Canada |
| **July** | |
| 16–19 | American Society of Civil Engineers, Duluth, Minnesota |

Journal of
Applied Physics

Volume 18, Number 5 May, 1947

# Special Issue on High Polymers

**The following six papers were presented at the meeting of the Division of High Polymer Physics of the American Physical Society at Columbia University, January 30–February 1, 1947.**

## Measurements of the Refractive Index of Films*

Fred W. Billmeyer, Jr.
*E. I. du Pont de Nemours and Company, Plastics Department, Arlington, New Jersey*

The refractive index of a film of transparent or translucent material may be determined by immersing the film in a suitable liquid mixture and observing the intensity of light reflected from the surface of the specimen. The composition of the immersion liquid is varied, and a record is made of its refractive index (measured on a refractometer) and of the reflected light intensity. The refractive index of the film is the same as that of the immersion liquid giving the lowest reflected intensity. This has been confirmed by experiments with a glass plate, whose refractive index was known. According to Fresnel's laws of reflection, the experimental data should fit a parabola with its minimum at the refractive index of the film. On fitting the data by the method of least squares, the minimum is obtained with a standard deviation of the order of 0.002 refractive index units. The method has been applied to several plastic materials.

### INTRODUCTION

A METHOD has been developed for measuring the refractive index of a solid material. Originally used with thin films of translucent plastics, it may be applied as well to thicker specimens, of more highly absorbing or colored material. Mentioned by R. W. Wood[1] and used to some extent in the investigation of anomalous dispersion phenomena, the procedure consists in immersing the specimen in a suitable mixture of liquids and observing the intensity of light reflected from its surface. The composition of the immersion liquid is varied, and a record is made of its refractive indices (measured by a refractometer) and of the corresponding intensities of the reflected light. The refractive index of the sample is the same as that of the immersion liquid which gives the lowest intensity of reflected light.

### THEORY

When a material of refractive index $n_0$, immersed in a liquid of refractive index $n$ not too far different from $n_0$, is illuminated with unpolarized light striking its surface at an angle $i$ (see Fig. 1), the fraction of the light reflected is given by Fresnel's Law as[2]

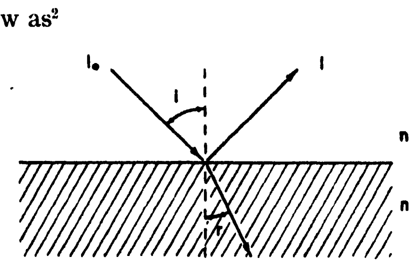


Fig. 1.

---

[*] Presented at the meeting of the Division of High Polymer Physics, American Physical Society, January 30–February 1, 1947.

[1] R. W. Wood, *Physical Optics* (The Macmillan Company, New York, 1934), third edition, pp. 411–12, 509.

[2] The coefficient ½ in these equations becomes nearly 1 when reflections from the back face of a transparent sample are included. For these experiments its actual value is immaterial.

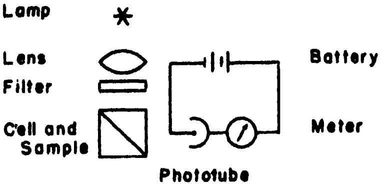


FIG. 2.

$$\frac{I}{I_0}=\frac{1}{2}\left[\frac{\sin^2(i-r)}{\sin^2(i+r)}+\frac{\tan^2(i-r)}{\tan^2(i+r)}\right], \qquad (1)$$

where $I_0$ and $I$ are the initial and reflected light intensities, respectively, and $i$ and $r$ are the angles of incidence and refraction. The angle $r$ is given by Snell's Law:

$$n \sin i = n_0 \sin r. \qquad (2)$$

If $i$ is chosen as 45°, a convenient angle experimentally, and $n$ is nearly equal to $n_0$, $r$ becomes nearly 45°, and Eq. (1) reduces to

$$I/I_0=\tfrac{1}{2}(\tan^2(i-r)-\tan^4(i-r)). \qquad (3)$$

Use of Eq. (2) then gives, for $i-r$ small,

$$I/I_0=\tfrac{1}{2}\left(\frac{n-n_0}{n_0}\right)^2+\cdots \qquad (4)$$

which is a parabola with its minimum at $n=n_0$.

## EXPERIMENTAL

### Apparatus

A diagram of the essentials of the apparatus is shown in Fig. 2.

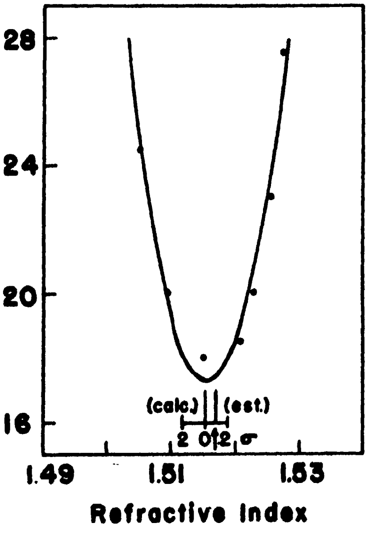


FIG. 3. Glass plate.

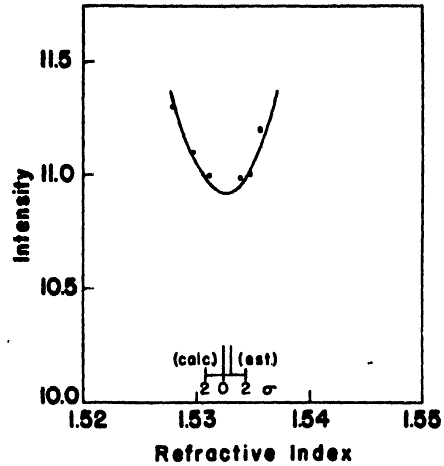


FIG. 4. Type FM-3 Nylon.

In the actual experiments two sets of apparatus were used. One consisted of a Bausch and Lomb projection microscope illuminator as light source and a Photovolt photometer, model 500, as detector. The tungsten lamp was run from a voltage-stabilizer, and a Wratten 62 filter was used to approximate mercury green light.

Later experiments were made with the light-scattering apparatus built by the author (to be described elsewhere). A stabilized incandescent lamp was used in conjunction with Wratten filters 11 and 22 to approximate sodium yellow light. Occasionally the filters were omitted.

In each case the sample was a thin film held between two pieces of polythene approximately 0.080 inch thick, arranged so as to fit tightly along the diagonal of a 4-centimeter square glass cell. The polythene was cut away so as to allow the light to hit the film and to allow free circulation of the immersion liquid. Polythene is especially suitable because of its general insolubility.

The immersion liquids must be miscible with one another and must not dissolve, swell, or otherwise affect the sample. Some liquid pairs which have been used are benzyl alcohol and carbon tetrachloride for the range in refractive index around 1.5, and carbon tetrachloride and benzene, or carbon tetrachloride and ethanol, for lower indices.

Measurements were carried out by starting with a liquid mixture about 0.01 refractive index unit away from the index of the film (chosen by experience or in a preliminary experiment) and adding small increments of a liquid with index beyond that of the sample. For each liquid mixture used, the refractive index and the intensity of reflected light were measured.

The measurements of refractive index were carried out either on a Bausch and Lomb "Precision" refractometer using mercury green light, or on a Bausch and Lomb Abbé refractometer, which gives a reading for sodium light. All experiments were carried out at room temperature (approximately 25°C) and the refractometers were checked against standard glass plates occasionally.

### Glass Plate Experiment

To test the accuracy of the procedure, the refractive index of a microscope slide was measured both by the method described here and directly by the Abbé refractometer, with the following result:

by refractometer,

$$n = 1.5171$$

by light reflection,

$n = 1.5156 \pm 0.0036$ (95 percent reliability).

The two values agree within one standard deviation. The data are presented in Fig. 3, with the refractometer value indicated by an arrow.

TABLE I. Refractive indices of films.

| Material | Index | Color of light | Notes |
|---|---|---|---|
| polythene | 1.515 | mercury green | PM-1, lot 720 |
|  | 1.519 | mercury green | PM-1, lot 859 |
| tetrafluoroethylene | 1.378 | white | "Teflon," density 2.18 |
|  | 1.376 | sodium yellow | "Teflon," density 2.12 |
| hexamethylene sebacamide | 1.532 | sodium yellow | Type FM-3 Nylon |

### COMPUTATIONS

The refractive index of the test specimen may be obtained from the experimental data either by inspection or from the equation of the parabola fitted to the observed points by the method of least squares.

In the latter case, the reflected intensity is represented by the equation

$$I = a + bn + cn^2 \qquad (5)$$

and the position of the minimum is

$$n_0 = -\frac{b}{2c} \qquad (6)$$

In the experiments carried out on films of plastics, the standard deviation in the value of $n_0$ was computed directly from the experimental data for the case of the glass plate. For each of the other cases the standard deviation was estimated from this value and the relative magni-

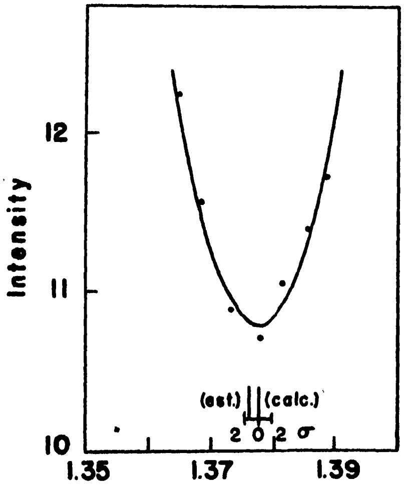


FIG. 5. "Teflon," density 2.18.

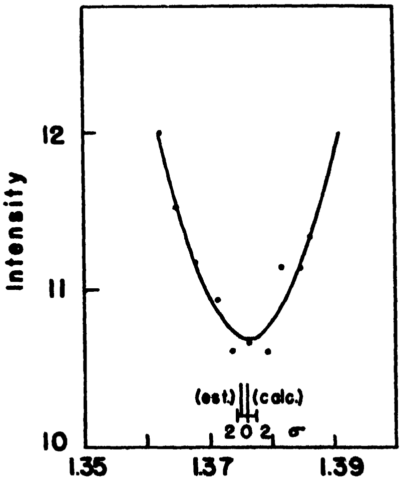


FIG. 6. "Teflon," density 2.12.

tudes of the "chi-square" index of precision[3] for glass and for the case in question.

Whenever the refractive index as determined by a visual inspection of the data was compared with that computed by least squares, the difference between the two was less than twice the standard deviation of the computed value.

Some experimental results are presented in the following table, and a few of the computed curves are shown in Figs. 4, 5, and 6. Included on the graphs are the estimated and computed indices of refraction of the respective films, and a scale of standard deviation, $\sigma$. The intensity is represented in arbitrary units, with the zero of intensity considerably below the axis.

The intensity at the minimum will depend upon the transparency of the film. For a very transparent material such as glass, it will be limited chiefly by the scattered and stray light in the apparatus. The value of $n_0$ is not influenced by the magnitude of this residual intensity.

## CONCLUSION

A method similar to the mineralogical immersion technique for measuring the refractive index of a solid has been presented. The test specimen may be a thick or thin piece with a reasonably plane face, or a film. The material may be clear or translucent, and may also be colored. The optical quality of its surface need only be fair. For these reasons this technique has proved successful in several cases where other methods, such as the determination of Brewster's angle, have failed.

Some experimental results are given in the accompanying table and figures.

---

[3] See, for example, R. A. Fisher, Statistical Methods for Research Workers (Oliver and Boyd, 1941), eighth edition.

---

# A Sound Velocity Method for Determination of Molecular Weight of Liquid Polymers†

Alfred Weissler, James W. Fitzgerald,* and Irving Resnick**
*Naval Research Laboratory, Washington, D. C.*

A new method is presented for determining number average molecular weights of liquid polymers, by means of easily performed measurements. The molecular weight is an explicit function

$$M=\frac{Bd}{v^{\frac{1}{3}}-A\left(\frac{n^2-1}{n^2+2}\right)}$$

of the sound velocity $v$, density $d$, refractive index $n$, and the two empirical constants $A$ and $B$. Accuracy of about 2 parts per hundred was attained for the lower polyethylene glycols. The method seems suitable for molecular weights up to several thousand. Sound velocities were measured by an acoustic interferometer, at one megacycle frequency.

The molecular weight of a polymer is so fundamental a property, and the methods available for its determination are so few in number, that any new technique warrants a certain interest. This paper presents such a new method, suitable for liquids, which is based on the relatively simple measurements of sound velocity, density, and refractive index.

## I. BASIS OF THE METHOD

Sound velocity in liquids became useful for the investigation of molecular structure after the discovery by M. R. Rao[1] of the empirical constant

$$R=Mv^{\frac{1}{3}}/d \qquad (1)$$

where $M$ is the molecular weight, $v$ and $d$ are the sound velocity and density measured at the same temperature, and $R$ may be called the molar sound velocity. For each pure liquid (except water) the molar sound velocity is a constant independent of temperature; therefore, density and the cube root of velocity vary in such a way that their ratio is fixed.

---

† Presented at the meeting of the Division of High Polymer Physics, American Physical Society, January 30–February 1, 1947.

* Now at Engineering Research Associates, Washington, D. C.

** Now at Johns Hopkins University, Baltimore, Maryland.

[1] M. R. Rao, Ind. J. Phys. 14, 109 (1940); J. Chem. Phys. 9, 682 (1941).

TABLE I. Physical properties of polyethylene glycols at 30°C.

| Glycol | Mol. wt. | Density | Sound velocity | $n_D$ | Viscosity | Boiling point |
|---|---|---|---|---|---|---|
| Ethylene | 62.1 | 1.1068 g/cc | 1642.9 m/s | 1.4295 | 13.95 c.p. | 69°/1.2 mm |
| Diethylene | 106.1 | 1.1094 | 1567.7 | 1.4438 | 22.55 | 105°/2.0 |
| Triethylene | 150.1 | 1.1158 | 1593.3 | 1.4531 | 29.96 | 123°/1.2 |
| Tetraethylene | 194.2 | 1.1159 | 1580.1 | 1.4567 | 35.16 | 138°/0.5 |
| Pentaethylene | 238.2 | 1.1165 | 1580.0 | 1.4593 | 42.37 | 176°/0.5 |
| Poly 200 | 200±10 | 1.1168 | 1592.2 | 1.4572 | 39.71 | |
| Poly 300 | 300±15 | 1.1176 | 1578.0 | 1.4618 | 59.02 | |
| Poly 400 | 400±20 | 1.1182 | 1576.0 | 1.4636 | 74.70 | |
| Poly 600 | 600±30 | 1.1183 | 1570.2 | 1.4653 | 104.62 | |

Although the velocity of sound varies irregularly within a given homologous series, such as the alcohols or ketones, the molar sound velocity was found to be an additive property of the atoms (or bonds) in the molecule. From the published tables of $R$ increments for atoms of carbon, hydrogen, oxygen, etc., the molar sound velocity may be calculated with good accuracy, just as in the familiar case of molar refraction.

Further, Lagemann and Dunbar[2] showed that, within a homologous series, linear relationships exist between any two of the following molar constants: (a) molar sound velocity, (b) molar refraction

$$N=\frac{M}{d}\left(\frac{n^2-1}{n^2+2}\right) \tag{2}$$

where $n$ is index of refraction, (c) molar viscosity as defined by Souders[3]

$$I=(2.9+\log\log\eta)\frac{M}{d} \tag{3}$$

where $\eta$ is viscosity in millipoises, (d) molar magnetic rotation, (e) parachor, (f) critical volume, and (g) van der Waals' $b$.

An example of such a relationship is the one involving refraction and sound velocity

$$R=AN+B \tag{4}$$

where $A$ and $B$ are, respectively, the slope and intercept. Upon substitution of Eqs. (1) and (2), this immediately yields an expression for the molecular weight

$$M=\frac{Bd}{v^{1/3}-A\left(\frac{n^2-1}{n^2+2}\right)} \tag{5}$$

in terms of the two empirical constants and the observed sound velocity, density, and refractive index. Similar expressions may be obtained from other pairs of molar constants; for example:

$$M=\frac{B'd}{v^{1/3}-A'(2.9+\log\log\eta)} \tag{6}$$

and

$$M=\frac{B''d}{\left(\frac{n^2-1}{n^2+2}\right)-A''(2.9+\log\log\eta)} \tag{7}$$

where $A'$, $A''$, $B'$, and $B''$ are the analogous slopes and intercepts of the other linear relations.

These results should be applicable not only to a homologous series, in which the difference between two successive members is $-CH_2-$, but also to a polymer homologous series, where the difference is any monomer structural unit.

## II. EXPERIMENTAL

It was therefore decided to test the validity of Eq. (5), and incidentally of (6) and (7), by using it to determine the molecular weights of several pure members and commercial mixtures of the polyethylene glycol series. These polymers, of type formula $HO-(-CH_2-CH_2-O-)_n-H$, have been studied intensively in the past.[4,5] The

---

[2] R. T. Lagemann and W. S. Dunbar, J. Phys. Chem. 49, 428 (1945).

[3] M. Souders, J. Am. Chem. Soc. 60, 154 (1938).

[4] H. Staudinger, *Der Aufbau der hochmolekularen organischen Verbindungen*, (Julius Springer, Berlin, 1932).

[5] R. Fordyce and H. Hibbert, J. Am. Chem. Soc. 61, 1910, 1912 (1939).

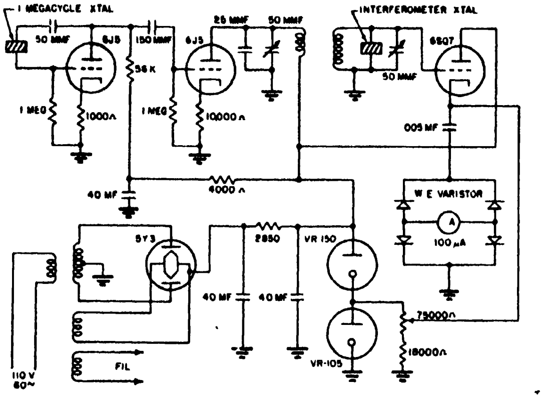


FIG. 1. Circuit design for interferometer.

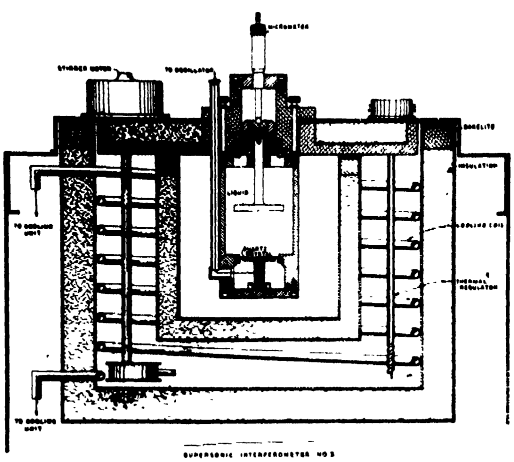


FIG. 2. Interferometer in thermoregulated oil bath. The inner tank shown has since been removed.

chemical individuals (the monomer, dimer, trimer, tetramer, and pentamer) were purified by fractional distillation through a 40 cm Widmer column, at about 1 mm pressure. The mixtures were those of average molecular weight 200, 300, 400, and 600, as manufactured by Carbide and Carbon Chemicals Corporation.

Several physical properties of these nine liquids were measured at 30°±0.1°: refractive index for sodium light by the Abbé refractometer, density by a 25 ml pycnometer, sound velocity by the acoustic interferometer described below, and viscosity by the Ostwald-Fenske viscometer. These properties are listed in Table I.

The acoustic interferometer[6] was of conventional type,[7-9] energized by a one megacycle crystal-controlled oscillator. Figure 1 shows the circuit used. Figure 2 is a sectional view of the brass interferometer cell, containing the quartz crystal source and the movable reflecting plate, immersed in the thermoregulated oil bath. About 300 ml of liquid are required for a measurement. As the reflector is moved vertically through the liquid by the micrometer head, cyclical variations in the standing wave pattern recur at distances of integral half-wave-lengths. These cause corresponding cycles in the current through the interferometer crystal, which are observed on a suitable microammeter. The micrometer-head travel required for an interval of two current maxima, then, is equal to the wave-length; and the velocity of sound is, of course, the product of this wave-length and the frequency. In practice, one measures the travel for twenty maxima, which makes possible a precision of a few hundredths of a percent.

## III. DISCUSSION

From the physical properties of the nine liquids, there were computed the molar refractions, viscosities, and sound velocities. These observed molar constants agree reasonably well (Table II) with the values calculated[10] from the molecular structures, using the published tables[2,3,10] of the bond increments for each type of constant. In the case of molar sound velocity, the agreement is better than 1 percent, which is less than the dispersion in values among the original com-

TABLE II. Molar constants of polyethylene glycols.

| | Molar sound velocity | | Molar refraction | | Molar viscosity | |
|---|---|---|---|---|---|---|
| Glycol | Observed | Calculated | Observed | Calculated | Observed | Calculated |
| Ethylene | 661.7 | 652.0 | 14.47 | 14.49 | 181.2 | 176.0 |
| Diethylene | 1111.2 | 1106.1 | 25.40 | 25.39 | 312.9 | 316.9 |
| Triethylene | 1571.6 | 1560.1 | 36.37 | 36.31 | 443.2 | 457.8 |
| Tetraethylene | 2027.1 | 2014.2 | 47.36 | 47.28 | 575.3 | 598.7 |
| Pentaethylene | 2485.1 | 2468.2 | 58.36 | 58.27 | 708.3 | 739.6 |
| Poly 200 | 2091 | 2075 | 48.79 | 48.87 | 593.6 | 617.5 |
| Poly 300 | 3125 | 3105 | 73.71 | 73.56 | 897.3 | 937.5 |
| Poly 400 | 4163 | 4135 | 98.63 | 98.48 | 1201.3 | 1257.5 |
| Poly 600 | 6236 | 6195 | 148.40 | 148.17 | 1813.4 | 1897.5 |

---

[6] Designed by one of us (J. W. F.) and Mr. Burton G. Hurdle.

[7] G. W. Pierce, Proc. Am. Acad. 60, 269 (1925).

[8] J. C. Hubbard and A. L. Loomis, Phil. Mag. 5, 1177 (1928).

[9] E. Klein and W. D. Hershberger, Phys. Rev. 37, 760 (1931).

[10] Molar refraction was calculated by the method of K. Fajans, *Physical Methods of Organic Chemistry*, edited by A. Weissberger (Interscience Publishers, New York, 1946), Vol. I, pp. 672, 679.

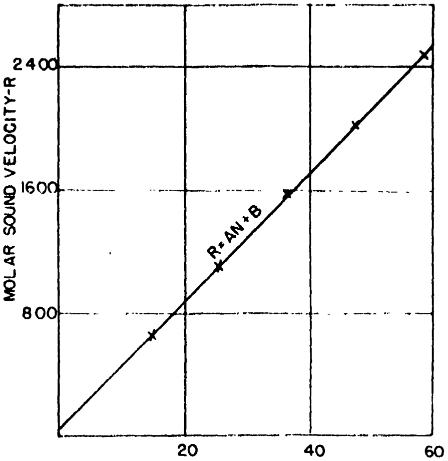


FIG. 3. Linear relation between molar sound velocity and molar refraction.

pounds used for constructing the table of increments. Compared to values previously reported[11] for 20°, the molar refractions show the customary increase of 0.01–0.02 percent per degree, attributed by Fajans[10] to loosening of the electron shell. The agreement is poorest for the molar viscosities.

Next, the existence of a linear relation between molar sound velocity and molar refraction was verified for the five pure compounds. The points lie on the straight line of Fig. 3, of which the slope $A$ is 41.59 and the intercept $B$ is 57.7. The slope may also be calculated from the tables of bond increments: for each $-CH_2-CH_2-O-$ structural unit, the $R$ increment is 454 and the $N$ increment is 10.88. The quotient of these two, which should be the slope, is 41.73, so that the agreement is a few parts per thousand.

TABLE III. Determination of molecular weight by sound velocity.

$$M=\frac{57.7d}{v^{\frac{1}{3}}-41.59\left(\frac{n^2-1}{n^2+2}\right)}$$

| Glycol | Theoretical Mole. wt. | Mole. wt. by sound velocity | Deviation | Mole. wt. number avg. | Mole. wt. wt. avg. |
|---|---|---|---|---|---|
| Ethylene | 62.1 | 59.9 | −3.5% | 60.8 | 62 |
| Diethylene | 106.1 | 111.5 | 5.1 | 101.6 | 109 |
| Triethylene | 150.1 | 147.3 | −1.9 | 148.5 | 155 |
| Tetraethylene | 194.2 | 195.7 | −0.8 | 195.4 | 190 |
| Pentaethylene | 238.2 | 237.7 | 0.2 | 234.5 | 238 |
| Poly 200 | 200±10 | 186.2 | −2.0+ | 183 | 219 |
| Poly 300 | 300±15 | 302.7 | 0.0+ | 315 | 359 |
| Poly 400 | 400±20 | 381.8 | 0.0+ | 384 | 480 |
| Poly 600 | 600±30 | 546.9 | −4.0+ | 561 | 720 |

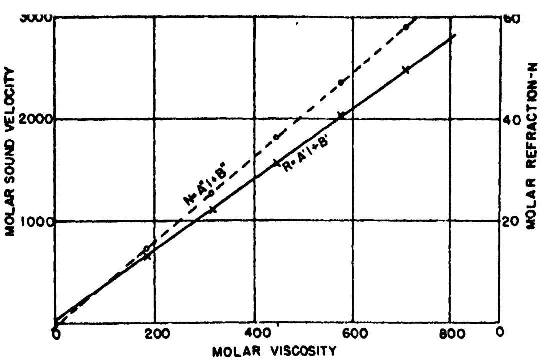


FIG. 4. Verification of other linear relations.

As expected, linear relations were found (Fig. 4) for the two other pairs of molar constants, sound velocity and viscosity, and refraction and viscosity. For the former pair, slope $A'=3.466$ and intercept $B'=31.9$; for the latter, $A''=0.08337$ and $B''=-0.640$. The agreement between observed and calculated slopes is not as good in these cases.

Next came the crux of the entire work. For each of the nine liquids, the molecular weight was calculated as indicated by Eq. (5), from the density, sound velocity, index of refraction, $A$, and $B$. The results show an average accuracy of about 2 percent (Table III) on the basis of the theoretical molecular weights, which compares well with that of other methods.

For purposes of comparison, the number average molecular weight was determined by freezing-point depression in water as solvent.[12] Measurements were made at concentrations of approximately 2 percent and 4 percent, and the values extrapolated to infinite dilution. The precision is estimated at about 3 parts per hundred. It is apparent from Table III that, within experimental error, the sound velocity molecular weight is the same as the number average

---

[11] A. F. Gallaugher and H. Hibbert, J. Am. Chem. Soc. **58**, 813 (1936).

[12] Using water as solvent avoided the anomalous results encountered in ethylene dibromide solution by Gallaugher and Hibbert, reference 11, presumably because the glycols tend to be solvated, rather than associated, in aqueous solution. The apparent molecular weights obtained in the present work decreased slightly with increasing concentration, which indicates that association of glycol molecules is negligible compared to solvation.

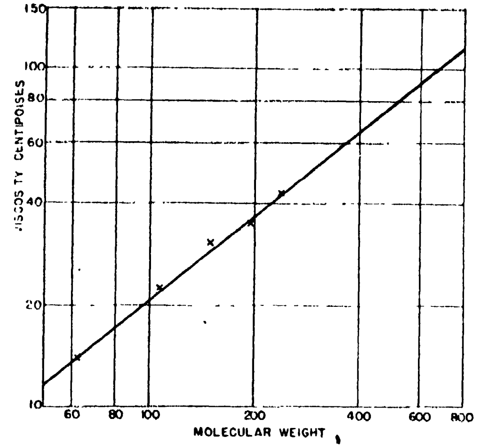


FIG. 5. Bulk viscosities of polyethylene glycols at 30°.

molecular weight. In addition, a weight average molecular weight was determined from the log-log plot of bulk viscosity against molecular weight[13] (Fig. 5). As would be expected, for the heterogeneous mixtures the weight average results shown in Table III are considerably higher than the number average or sound velocity molecular weights.

Attempts were also made to calculate the molecular weights from Eqs. (6) and (7), the analogous relations involving the other two pairs of properties. However, the errors are large, as demonstrated in Table IV. This may be due partly to the fact that the precision of measurement of viscosity is only about one-tenth as good as that of sound velocity or refractive index. Unfortunately, the errors increase for higher molecular weights because the calculations involve a decreasing difference between the two terms in the denominator of the expression for $M$.

## IV. SUMMARY

A new method for determining molecular weights of polymeric liquids has been shown to yield results of about 2 percent accuracy in the case of the lower members of the polyethylene glycol series. This method, which gives the number average molecular weight, consists of calculating $M$ from two empirical constants and the observed sound velocity, density, and refractive index. We intend to investigate extension of its applicability to solid polymers or solutions of solids.

## V. ACKNOWLEDGMENTS

Grateful acknowledgments are made to the many physicists and chemists at the Naval Research Laboratory who were generous with their facilities, materials, and advice during the prosecution of this work; and also to the Dow Chemical Company and to the Carbide and Carbon Chemicals Corporation, for generously supplying samples of these polyethylene glycols.

TABLE IV. Molecular weight determinations using other pairs of molar constants.

| Glycol | Theoretical Mol. wt. | Refraction and velocity | Velocity and viscosity | Refraction and viscosity |
|---|---|---|---|---|
| Ethylene | 62.1 | 59.9 | 63.6 | 62.5 |
| Diethylene | 106.1 | 111.5 | 131.4 | 98.2 |
| Triethylene | 150.1 | 147.3 | 138.0 | 167.2 |
| Tetraethylene | 194.2 | 195.7 | 186.0 | 205.2 |
| Pentaethylene | 238.2 | 237.7 | 237.9 | 221.9 |
| Poly 200 | 200±10 | 186.2 | 187 | 183 |
| Poly 300 | 300±15 | 302.7 | 479 | 185 |
| Poly 400 | 400±20 | 381.8 | 1677 | 168 |
| Poly 600 | 600±30 | -546.9 | Infinite | 137 |

---

[13] These lower polyethylene glycols appear not to conform to the relation $\log\eta = A + BM^{\frac{1}{2}}$ suggested for bulk viscosities by P. J. Flory, J. Am. Chem. Soc. **62**, 1057 (1940).

# An Instrument for Measuring Particle Diameters and Constructing Histograms from Electron Micrographs*

E. E. HANSON AND J. H. DANIEL
*Chemical and Physical Research Laboratories, The Firestone Tire and Rubber Company, Akron, Ohio*

An instrument for shortening the time and labor required in constructing particle size histograms from electron micrographs is described. Representative histograms obtained with the instrument for GR-S latices are shown. Linear, surface, and volume-equivalent diameters of non-spherical particles are defined, and variations in the instrument for plotting such diameters are proposed.

## INTRODUCTION

IN certain studies of the polymerization of synthetic rubbers, it is desirable to obtain particle-diameter distribution curves or histograms for the rubber latices. Such diameter histograms may be obtained from electron microscope photographs of the latex particles.

The technique used in this laboratory for preparing latex specimens for the electron microscope[1] consists essentially of forming a film from a dilute water solution of polyvinyl alcohol containing a small amount of the latex.

The construction of particle diameter histograms is tedious and time consuming. In the usual procedure, the "longest" and "shortest" diameters of the projected image (or of a photographic enlargement) of a particle are measured with a rule, their arithmetic mean is computed, and finally the number of particles with mean diameters lying in chosen diameter intervals is plotted against the diameters. Multiplication by a scale factor, containing the reciprocal of the magnification, converts the diameter axis of the histogram to read directly in Angstrom units. To obtain a representative distribution, at least 200 particles should be measured.

The instrument to be described here was developed to reduce the time and labor required in constructing such histograms. It will be called a "histogram computer."

In the development of the histogram computer, consideration was given to the question of whether the arithmetic or the geometric mean, or some other function of the "longest" and "shortest" diameters of a particle, should be plotted. With an unbroken film in the electron microscope, the latex particles usually appear approximately spherical in shape. However, when holes are present in the film, as frequently is the case, the particle contours are elliptical. The problem, then, is to compute the diameter of the equivalent spherical particle having either the same volume, or the same surface area, as the actual ellipsoidal particle. The two cases require, of course, calculation of different functions of the major and minor axes. In addition, the

TABLE I. Comparison of arithmetic and geometric mean diameters with volume and surface equivalent diameters.

| | | | | Prolate Spheroid | | | | | | Oblate Spheroid | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | % Error | | | % Error | | | % Error | | | % Error | |
| Ratio $a/b$ | Arith. Mean $\frac{a+b}{2}$ | Geom. Mean $(ab)^{\frac{1}{2}}$ | Difference % | $D_v$ | Arith. Mean | Geom. Mean | $D_s$ | Arith. Mean | Geom. Mean | $D_v$ | Arith. Mean | Geom. Mean | $D_s$ | Arith. Mean | Geom. Mean |
| 1.0 | $b$ | $b$ | 0.0 | $b$ | 0.0 | 0.0 | $b$ | 0.0 | 0.0 | $b$ | 0.0 | 0.0 | $b$ | 0.0 | 0.0 |
| 1.5 | $1.25b$ | $1.225b$ | 1.0 | $1.145b$ | 9.2 | 7.0 | $1.16b$ | 7.8 | 5.6 | $1.31b$ | −4.6 | −6.5 | $1.33b$ | −6.0 | −7.9 |
| 2.0 | $1.50b$ | $1.414b$ | 2.9 | $1.260b$ | 19 | 12.2 | $1.31b$ | 14 | 7.9 | $1.587b$ | −5.5 | −11 | $1.66b$ | −9.6 | −15 |
| 2.5 | $1.75b$ | $1.58b$ | 5.1 | $1.357b$ | 29 | 16 | $1.44b$ | 22 | 9.7 | $1.842b$ | −5.0 | −14 | $1.99b$ | −12 | −21 |

$a$ = Major Axis
$b$ = Minor Axis

$D_v$ = Diameter of sphere of equal volume.
$D_s$ = Diameter of sphere of equal surface.

---

* Presented at the meeting of the Division of High Polymer Physics, American Physical Society, New York, January 30–February 1, 1947.

[1] R. H. Kelsey and E. E. Hanson, J. App. Phys. 17, 675 (1946).

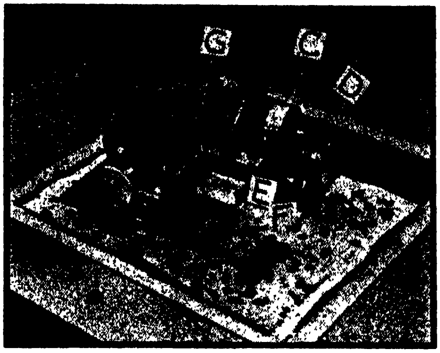


FIG. 1. The scanning mechanism.

particles may be either prolate or oblate, or some shape in between. Without going through some time-consuming process, such as taking stereo-photographs, there is no way of determining what volumetric shape a particle has.

A comparison between the diameters of the volume-equivalent and the surface-equivalent spheres, and the arithmetic and geometric mean diameters for both prolate and oblate spheroids is given in Table I. In some cases the geometric mean is the closer approximation, and in others the arithmetic mean is the closer. It was therefore decided to design the simplest of possible mechanisms, which gives the arithmetic mean. In general, the measurements should be restricted to particles having a major to minor axis ratio of less than two.

## APPARATUS

The histogram computer consists essentially of a scanning mechanism for setting manually two pairs of mutually perpendicular lines on the outline of the particle, and a coupled recording mechanism for dropping a steel ball into a compartment corresponding to a range of diameters containing the arithmetic average of the two measured diameters of the particles.

Figures 1 and 2 are photographs of the scanning and the recording mechanisms, respectively. The scanning mechanism consists of a steel baseplate with a right angle section cut out for viewing the electron micrograph field, two mutually perpendicular screws (one labelled $E$) for translating non-rotating nuts to which clear Lucite blocks are attached, two equal gear trains coupling the rotations of the screws to their respective Selsyn motors, and a sub-base plate of clear Lucite on the under side of which are two scribed, inked-in lines drawn parallel with, and perpendicular to the two screws. Each of the movable Lucite blocks has a scribed, inked-in line drawn on its top and bottom surface, in a vertical plane parallel to a line in the sub-base plate. This enables the operator to set the lines without parallax on the edge of the particle. Each of the two Selsyn motors is coupled to a corresponding motor on the recorder so as to transmit its rotation.

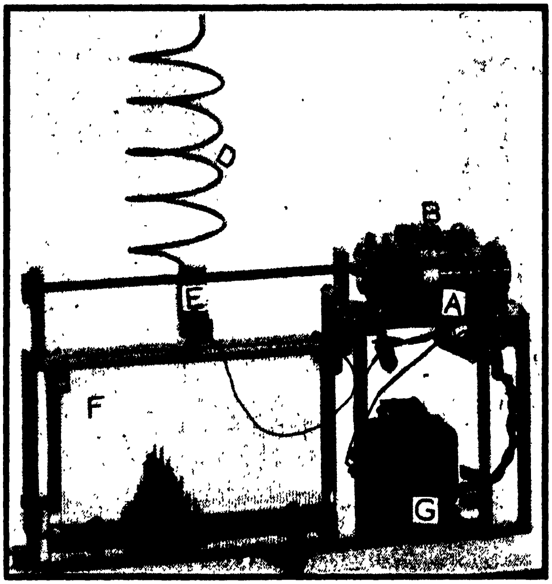


FIG. 2. The recording mechanism.

In the recorder unit, the Selsyn motors, $A$ and $B$, transmit their rotations additively through a differential gear, $C$, to a screw 10 inches long. The screw serves to position the ball-dispensing mechanism $E$. Directly below the ball-delivery

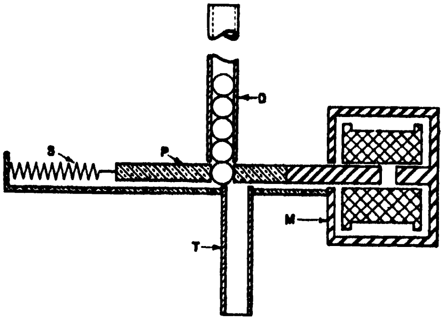


FIG. 3. The ball release mechanism.

tube, and parallel to the screw, is a thin rectangular Lucite box, *F*, divided into 80 compartments for receiving the steel balls as they are dropped.

The ball-dispensing unit consists of a $\frac{1}{8}''$ I.D. copper-tubing magazine, *D*, for holding a supply of $\frac{3}{32}''$ diameter steel balls, and an electromagnetically controlled "valve" by which the operator can cause the balls to be dropped one at a time. A cross-sectional diagram of the valve is shown in Fig. 3. The axis of the delivery-tube is sufficiently offset from the magazine, *D*, to prevent the balls from falling through. The flat plunger, *P*, normally held by the spring, *S*, in the position shown, is pierced by a hole slightly larger in diameter than the steel balls. When the circuit to the magnet, *M*, is closed, the plunger carries one ball over and discharges it into the delivery tube. The delivery tube projects between two parallel guides to a position $\frac{3}{16}''$ above the opening in the recorder compartments. The parallel guides also serve the purpose of preventing any rotation of the dispensing unit nut.

The recorder compartment box, *F*, was made by bolting together the two ends of two pieces of $6''\times11''\times\frac{1}{4}''$ Lucite blocks against .096″ thick spacers. Strips of .010″ thick shim stock were inserted into grooves which had been accurately milled into the inner surfaces of the plates on $\frac{1}{8}''$ centers. The bottom of the box is a metal blade which can be removed at the end of a run to release the accumulated balls into a flat box below.

End-play was taken out on the three screws by special thrust bearings. Back-lash was eliminated through the use of split-nuts engaging the screws.

The Selsyn motors were designed to be operated normally at 115 volts and 400 cycles. However, since no 400 cycle source was available, they were operated from the transformer, *G*, at 60 cycles and 24 volts.

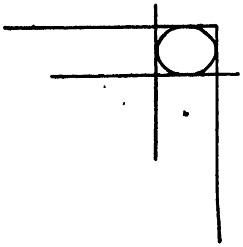


FIG. 4. Appearance of the scanning lines set on a particle.

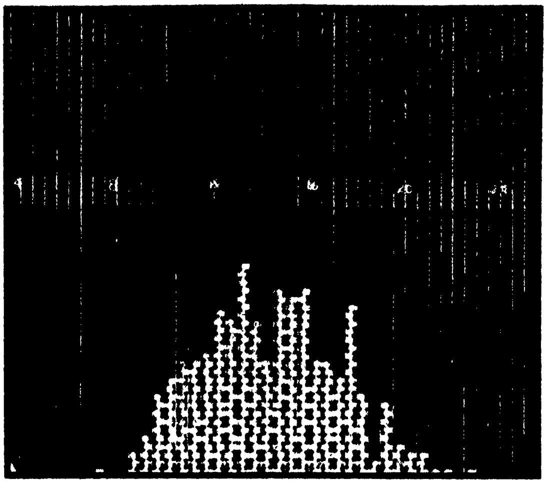


FIG. 5. Contact print of the recorder compartmented box showing distribution from which histogram of run 3, Fig. 6, was obtained.

The gear train normally used gave an amplification factor of 18 from the scanning to the recorder units. Therefore, a total magnification of 35,300 diameters was required in the electron micrograph to give a value of 50 Angstroms to each $\frac{1}{8}''$-long compartment interval. Other values could be given to the compartment interval by a suitable choice of gears and total magnification of the electron micrograph.

In practice, a particle count can be made either from a photographic enlargement, or directly from an image of the electron micrograph plate projected at the desired magnification on a ground-glass screen. The procedure is first to set the scanner unit on the zero position (scribed lines directly over each other), and the ball dispensing unit approximately at the zero position of the recorder box, and finally to shift the compartmented box slightly by hand until the balls fall equally often on both sides of the chosen zero wall. The scanning unit is then set successively on the particles to be measured, so that each particle is bracketed by the scribed lines as shown in Fig. 4. As each particle in turn is thus bracketed, the ball release switch is closed to record that particle. A counter, coupled to the switch, automatically totals the number of particles as the measurements proceed.

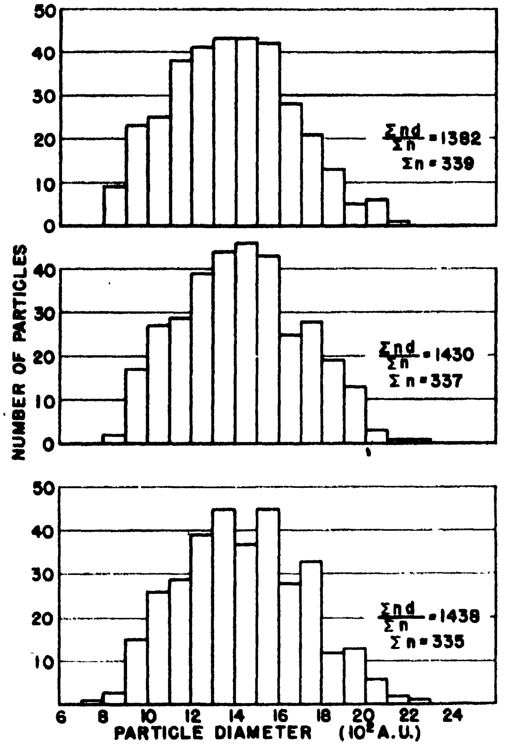


FIG. 6. Histograms of experimental latex. Each run made from same electron micrograph field.

### TYPICAL RESULTS

Figure 5 is a direct contact print of the record given by the histogram computer of 335 particles of an experimental GR-S latex. The print was made by placing F-2 contact paper against the back of the recorder compartments, and illuminating the front of the box for 15 seconds with a 60 watt Mazda bulb placed 7 feet away. In this case the electron microscope magnification was 5600, and the photographic enlargement was 6.3× to give a total magnification of 35,300×.

In Fig. 6 are three histograms of an experimental GR-S latex. All were constructed from the same electron micrograph field, and are, therefore, indicative of the reproducibility of the instrument. In these measurements an interval of 100 Angstrom units was chosen. Since the limit of resolution of the electron microscope was about 50 Angstroms for the specimens used, it was concluded that the sharp variations between 50 Angstrom intervals, shown in Fig. 5, were not significant.

Figure 7 is similarly a record of three successive histograms constructed from the same field of an electron micrograph of another GR-S type latex.

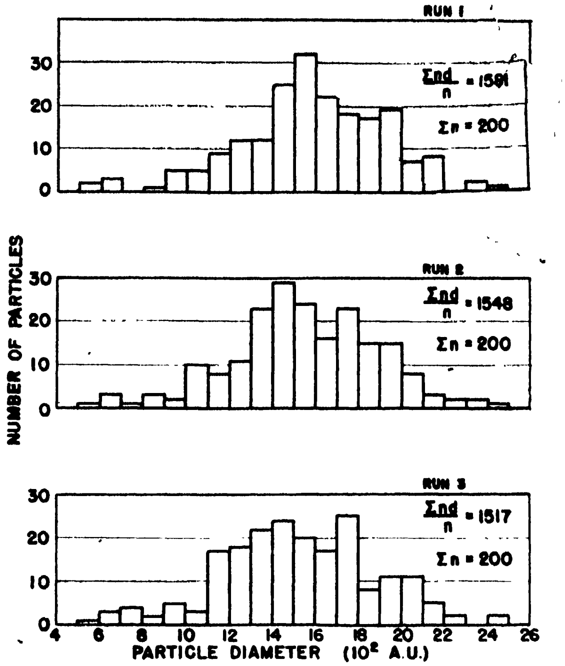


FIG. 7. Histograms of another experimental latex. Each run made from same electron micrograph field.

In general, the reproducibility of the curves is considered satisfactory.

It has been our experience that it is much less tedious to operate the computer than to measure particles with a rule, that the accuracy is increased, and that the time required to construct a histogram is reduced by about one-half.

### OTHER POSSIBLE MECHANISMS

Construction of an electronic servo-system, to replace the Selsyns and differential in the histogram computer, was not justified in this laboratory because of extra complications, but would give added features which might be desirable in some applications. Among such features may be listed the following:

1. Distribution of the geometric as well as arithmetic means of major and minor axes.
2. Distribution of surface area (approximate) or of volume (exact), assuming particle shape to be an ellipsoid of revolution.
3. Flexibility of size of histogram increments.

One such scheme is shown in Fig. 8. Circular potentiometers, $P_1$ and $P_2$, have resistances linear with their respective angles of rotation, $\theta_1$ and $\theta_2$, to the desired degree of accuracy. In use, $\theta_1$ and $\theta_2$ are set proportional to the major and minor axes, $D_1$ and $D_2$, respectively. Similar potentiometers, $P_3$ and $P_4$, are mechanically inter-connected so that their angles of rotation remain always equal to the single angle $\theta$. This angle $\theta$ is determined by the position of the servo-motor, which also runs the ball dispenser of the recorder (minus differential gear). In Fig. 8, the two voltages, $V_2$ and $V_3$, applied to the sum amplifier will always differ 180° in phase, so when their magnitudes are equal, the sum amplifier output to the servo-motor will be zero. The direction of rotation of the servo-motor when $V_2$ and $V_3$ are not equal will depend on which voltage is the greater. The blocks labeled *CF* are cathode followers with amplification, $A_{CF}$, extremely constant and slightly less than unity. Setting $R_1 = R_2 = 0$ for the moment, the following relations will exist:

$$V_2 = V_1 A_{CF_1}\theta_2 = V_1 A_{CF_1} k_2 D_2$$

$$= k_1 D_1 C_1 V_0 A_{CF_1} k_2 D_2, \quad (1)$$

$$V_3 = V_4 A_{CF_2}\theta = -A_{CF_2}\theta^2 C_4 V_0, \quad (2)$$

where $\theta_1 = k_1 D_1$; $\theta_2 = k_2 D_2$; voltage across $P_1 = C_1 V_0$; voltage across $P_4 = -C_4 V_0$.

When the servo-motor has equalized $V_3$ and $-V_2$, and the recorder ball dispenser comes to rest, the following relation will exist:

$$A_{CF_2}\theta^2 C_4 V_0 = k_1 D_1 C_1 V_0 A_{CF_1} k_2 D_2,$$

$$\theta = \left(k_1 k_2 \frac{C_1}{C_4}\frac{A_{CF_1}}{A_{CF_2}}\right)^{\frac{1}{2}} (D_1 D_2)^{\frac{1}{2}}. \quad (3)$$

Hence the position of the recorder ball dispenser is proportional to the geometric mean of the particle major and minor axes.

Since $V_0$ does not appear in Eq. (3), line voltage regulation is unnecessary. Rheostat $R_1$ allows the increment of diameter for the histogram to be varied. The combination of $R_1$ and $R_2$ allows a shift of recorder zero for cases where no small particles are present.

To obtain a histogram of approximate surface areas, a switching arrangement feeds $V_4$ instead of $V_3$ to the sum amplifier, thus giving $\theta$ proportional to the product $D_1 D_2$.

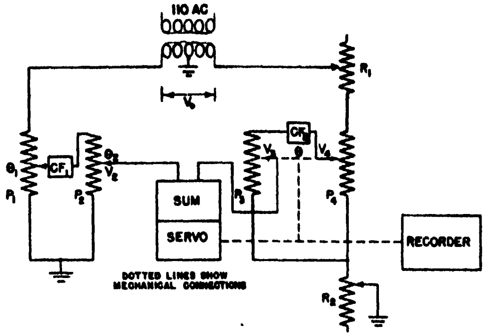


FIG. 8. Schematic diagram of proposed servo-mechanism for histogram computer.

To obtain a histogram of the arithmetic mean of particle diameters, a switching arrangement removes $CF_1$ and $CF_2$ from the circuit, connects the ungrounded ends of $P_1$ and $P_2$, and feeds $V_1$ and $V_2$ separately, and $V_4$ instead of $V_3$, to the sum amplifier (for this arrangement $k_1$ must equal $k_2$, as would normally be the case).

To obtain a histogram of volumes, $V_4$ instead of $V_3$ is fed to the sum amplifier, $P_3$ is connected mechanically and electrically to $P_2$ in the same manner it was formerly connected to $P_4$, and the series outputs of $P_1$, $P_2$, and $P_3$, are fed to the sum amplifier. In the case of a prolate spheroid, $\theta_1$ must be set proportional to the major axis; for an oblate spheroid, $\theta_1$ must be set proportional to the minor axis. Obviously, only a small range of particle sizes can be covered in volume histograms as compared to diameter histograms.

## ACKNOWLEDGMENTS

It is a pleasure to acknowledge the assistance given by Mr. George Cervana in laying out the design of the instrument and in operating it. Mr. F. S. Grover and Mr. Howard Hentzell did the constructional work and contributed valuable ideas to the design. We are indebted to Mr. G. E. White of the Sperry Gyroscope Company, Inc., for the gift of the small differential gear used in the recorder. Thanks are due the Firestone Tire & Rubber Company for permission to publish this paper, and to Dr. J. H. Dillon and Dr. F. W. Stavely for their interest in the problem.

# Torsion of a Rubber Cylinder*

R. S. Rivlin
*British Rubber Producers' Research Association, Welwyn Garden City, Herts., England, and National Bureau of Standards, Washington, D. C., U. S. A.*

It has been predicted theoretically that, in general, a right-circular cylinder of incompressible, highly elastic material, which is isotropic in its undeformed state, cannot be held in a state of pure torsional deformation by means of a torsional couple alone. In addition, normal surface tractions must be exerted over the plane ends of the cylinder. These normal surface tractions depend on the amount of torsion and on position on the plane ends of the cylinder. Experiments are reported here in which this phenomenon is observed in a right-circular cylinder of pure gum compound. The dependence of the surface traction on amount of torsion and its distribution over the surface of the cylinder is studied by measuring the bulging of the rubber into small holes in a metal plate on one end of the cylinder.

## I. INTRODUCTION

In a series of papers which is pending publication elsewhere,[1] the mechanical properties of highly elastic materials, such as rubber, which can be assumed incompressible and isotropic in their undeformed state, are discussed in terms of a stored-energy function $W$. This stored-energy function is specified in terms of the invariants of strain. It is shown that, once $W$ is specified in this manner, the forces, which must be applied in order to support any given state of deformation, can be calculated. This calculation has been carried out for the pure torsional deformation of a right circular cylinder, for an assumed stored-energy function of simplest possible type and subsequently for any general form of stored-energy function.

The theory predicts that in order to produce a pure torsion in the cylinder, a torsional couple must be exerted on the plane ends of the cylinder, as in the classical theory of elasticity for small deformations. In addition, normal tractions must be exerted over these surfaces. The distribution of these normal tractions over the ends of the cylinder and their dependence on the amount of torsion have been calculated for any specified form of the stored-energy function.

In this paper experiments are reported in which this distribution and dependence on amount of torsion is measured for a cylinder of pure gum compound, on the assumption that the surface traction is measurable by the bulging of the rubber into small holes in a metal plate on one end of the cylinder. The results obtained are compared with those derived theoretically for two simple forms of the assumed stored-energy function. Agreement, within the accuracy of the experiments, is obtained if the stored-energy function has a form equivalent to that put forward for rubber by Mooney.[2] This form involves two physical constants for the material.

## II. SPECIFICATION OF THE STORED-ENERGY FUNCTION

The stored-energy function for an ideal perfectly elastic material may be completely defined in the following manner. If a cube of the material, of unit edge, is subjected to a pure, homogeneous deformation in which it is deformed into a cuboid of dimensions $\lambda_1$ by $\lambda_2$ by $\lambda_3$, then the strain invariants $I_1$, $I_2$, and $I_3$ are defined as

$$I_1=\lambda_1^2+\lambda_2^2+\lambda_3^2,$$
$$I_2=\lambda_2^2\lambda_3^2+\lambda_3^2\lambda_1^2+\lambda_1^2\lambda_2^2$$

and

$$I_3=\lambda_1^2\lambda_2^2\lambda_3^2. \tag{1}$$

If the material is isotropic in its undeformed state, the energy stored in the deformed cube is defined as the stored-energy function for the material. It must be a function of $I_1$, $I_2$, and $I_3$, provided it depends only on the state of deformation of the cube and not on the manner in which the deformation is reached. Once the form of this

---

* Presented at the meeting of the Division of High Polymer Physics, American Physical Society, January 30–February 1, 1947.

[1] R. S. Rivlin, Phil. Trans. Roy. Soc. A (pending publication).

[2] M. Mooney, J. App. Phys. 11, 582 (1940).

function is specified, the mechanical properties of any body of this material are, in principle, completely determinate and are purely a matter for calculation, provided such phenomena as internal friction and creep are negligible.

For an incompressible material, the volume of the cube considered is unaltered in the deformation, so that

$$I_3=\lambda_1^2\lambda_2^2\lambda_3^2=1. \tag{2}$$

Then $W$, the stored-energy function, is given by

$$W=W(I_1, I_2). \tag{3}$$

In general, the form of the function $W(I_1, I_2)$ will be such that when the body is undeformed $W=0$.

### III. THE TORSION OF A RIGHT-CIRCULAR CYLINDER

Let us consider a right-circular cylinder of radius $a$ and length $l$ of incompressible material, for which the stored-energy function $W$ is given by (3). It has been shown that, in such a body, a pure torsion, in which each circular cross section of the cylinder is turned in its own plane through an angle $\phi(=\psi z)$ proportional to its distance $z$ from one end, can be maintained by the application of surface tractions alone. The necessary system of surface tractions is indeterminate to the extent of an arbitrary hydrostatic pressure, but if the surface tractions applied to the curved surface of the cylinder are zero, then the surface tractions which must be applied to the plane ends have been calculated as

(*i*) tangential surface tractions $\Theta$ acting azimuthally in the deformed state of the cylinder (these effectively act as a torsional couple on the cylinder), and
(*ii*) normal surface tractions $Z$.

$\Theta$ and $Z$ are measured per unit area of the surfaces on which they act and their distribution over the plane ends of the cylinder and dependence on $\psi$ are given by

$$\Theta=2\psi r\left(\frac{\partial W}{\partial I_1}+\frac{\partial W}{\partial I_2}\right)$$

and

$$Z=-2\psi^2a^2\frac{\partial W}{\partial I_2}+2\psi^2\int_a^r r\left\{\left(\frac{\partial W}{\partial I_1}-2\frac{\partial W}{\partial I_2}\right)-2\left[\frac{\partial^2 W}{\partial I_1^2}+(3+\psi^2r^2)\frac{\partial^2 W}{\partial I_1\partial I_2}+(2+\psi^2r^2)\frac{\partial^2 W}{\partial I_2^2}\right]\right\}\cdot dr. \tag{4}$$

$r$ is the distance of the point considered from the axis of the cylinder. $\partial W/\partial I_1$, $\partial W/\partial I_2$, $\partial^2 W/\partial I_1^2$, etc., are, in general, functions of $I_1$ and $I_2$ which can be calculated if the form of $W(I_1, I_2)$ is known. For the deformation considered,

$$I_1=I_2=3+\psi^2r^2. \tag{5}$$

We shall consider here two particular choices of the form of the stored-energy function $W$ for which Eq. (4) are considerably simplified.

In the series of papers[1] mentioned above, particular attention was paid to the stored-energy function given by

$$W=C_1(I_1-3), \tag{6}$$

when $C_1$ is a physical constant characterizing the material. Materials for which Eq. (6) are valid were described as incompressible, neo-Hookean materials.

The particular interest residing in this choice of the stored-energy function rests on the facts that

(*i*) it is the simplest choice that can be made, from the point of view of the formulation of a theory of large elastic deformations;
(*ii*) it is the form derived by Treloar,[3] for an ideal rubber, on the basis of the kinetic theory of large elastic deformations.

Equation (6) is a particular form, obtained by making $C_2=0$, of the relation

$$W=C_1(I_1-3)+C_2(I_2-3), \tag{7}$$

which was first suggested by Mooney[2] as suitable for describing the elastic properties of rubber. It is noted that Eq. (7) involves two physical constants for the material, $C_1$ and $C_2$.

If the relation (7) is introduced into (4), we

---

[3] L. R. G. Treloar, Trans. Faraday Soc. 39, 241 (1943).

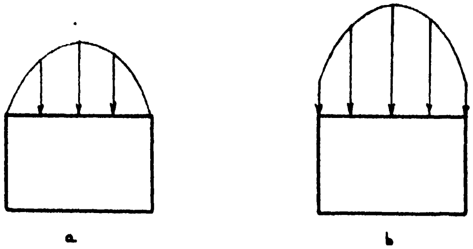


Fig. 1. Distribution of normal surface traction over plane ends of cylinder in pure torsion.
(a) When $W = C_1(I_1 - 3)$
and
(b) when $W = C_1(I_1 - 3) + C_2(I_2 - 3)$.

obtain

$$\theta = 2\psi r(C_1 + C_2)$$

and

$$Z = -2\psi^2 a^2 C_2 - \psi^2 (a^2 - r^2)(C_1 - 2C_2). \quad (8)$$

The equations for $\Theta$ and $Z$, when $W$ takes the form (6) can, of course, be obtained by putting $C_2 = 0$ in (8).

We see that if the physical properties of the material are represented by either Eqs. (6) or (7), the force $Z$ which must be exerted at each point is proportional to the square of the amount of torsion. If Eq. (6) applies, $Z$ must everywhere have the nature of a thrust (since $C_1$ is essentially positive), which is distributed over the diameter of a plane end as shown in Fig. 1(a). If Eq. (7) applies, then $Z$ does not vanish at the periphery (where $r = a$) and provided $C_2$ is positive and sufficiently small compared with $C_1$, its distribution will have the form shown in Fig. 1(b).

### IV. EXPERIMENTAL PROCEDURE

The experiments which will be described in this paper have as their object the study of the distribution of the normal force $Z$ defined in the previous section and its dependence on the amount of torsion.

In a preliminary qualitative experiment a rubber cylinder of pure gum compound, with brass plates bonded onto its plane ends, was subjected to a torsional couple by means of a lever attached to one of the ends while the other end was held in a vise. It was found that unless a longitudinal thrust was intentionally exerted parallel to the axis of the cylinder, the cylinder elongated, its axial cross section before and after torsion being shown roughly in Fig. 2. This simple experiment bears out the mathematical conclusion that a pure torsional couple does not give rise to a pure torsion. Clearly, if a pure torsion is to be obtained, normal compressive forces, whose resultant must have the nature of a thrust, must be exerted over the plane ends of the cylinder. This is in general agreement with the theoretical predictions outlined in the previous section.

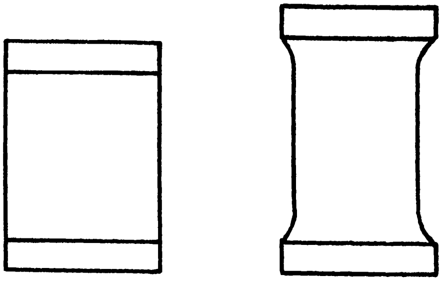


Fig. 2. Axial cross section of cylinder before and after twisting by a couple.

Consequently, an apparatus was designed in which the pure gum cylinder, to whose plane ends brass disks are bonded, is restrained from elongating. So, when one disk is rotated with respect to the other by the application of a torsional couple, the longitudinal thrusts required for the production of a pure torsion in the rubber cylinder are automatically called into play. The upper brass disk $A$ contains five circular holes, each ½-inch in diameter, varying in distance from the center and spaced as shown in Fig. 3, in order to have as little mutual interaction as possible. The lower disk $B$ has rigidly attached to it a steel disk $C$.

The assembly of rubber cylinder and disks $A$, $B$, and $C$ is held in a robust steel frame $D$, as shown in Fig. 4. The upper disk is rigidly attached to the frame and the lower disk is capable of being rotated in its own plane about its axis, by means of a crowbar which is introduced into one of a series of holes in the steel disk $C$. Provision is made for holding the cylinder in a state of torsion by means of steel pegs, which can be introduced into one of the holes $E$ in the frame. A scale is also provided for measuring the angle through which the lower disk is rotated.

When this disk is rotated, the rubber bulges slightly in the holes in disk $A$ and the amount of this bulging is assumed to be proportional to the thrust which would have to be exerted to prevent the bulging.

The bulging represents a small deformation superposed on a relatively large torsional deformation. At each of the holes this torsional deformation approximates to a state of simple shear of an amount depending on the amount of the torsion and the position of the hole considered. The assumption that the amount of bulging is proportional to the normal force tending to produce it is equivalent to an assumption that the former does not depend greatly on the state of simple shear on which the bulging deformation is superposed. There are certain theoretical reasons for assuming that this is approximately valid in the experiments described here.

In order to produce a simple extension in a cylinder superposed on the simple torsion, an additional normal surface traction must be exerted over each of the plane ends. If the stored-energy function for the material of the cylinder has the form (6), it has been shown[1] that this additional normal surface traction is independent of the amount of torsion and of position on the surface of the cylinder. The stress distribution in the neighborhood of the hole is complex, but for small deformations, it is reasonable to assume that the amount of bulge in each hole is dependent only on the force tending to produce the bulging and not on the amount of torsion or position of the hole, except insofar as this force is determined by them. This conclusion should be approximately valid if the stored-energy function has the form (7), provided $C_2$ is small compared with $C_1$. However, there is no reason to believe that it should be true for every possible form of the stored-energy function.

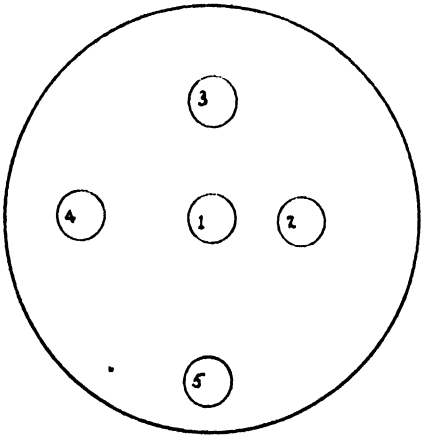


FIG. 3. Upper brass disk vulcanized to rubber cylinder.

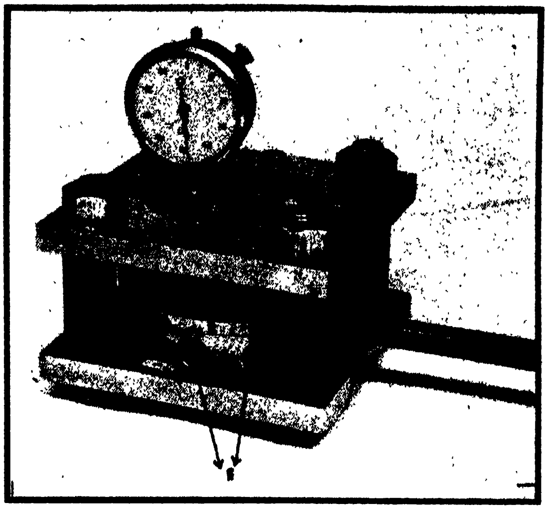


FIG. 4. The torsion apparatus.

The amount $d$ of the bulging in each of the holes is measured at the center of the hole, for various angles of rotation $\theta$ of the disk $B$, by means of a dial gauge. The top of the disk $B$ is ground flat and used as a reference plane in carrying out this measurement.

## V. EXPERIMENTAL RESULTS

If Eq. (8) is valid for $Z$, then the relation between $d$ and $\theta$ should be

$$d = K\theta^2[2a^2C_2 + (a^2 - r^2)(C_1 - 2C_2)], \qquad (9)$$

where $K$ is some constant of proportionality. $r$ is now taken as the distance of the center of the hole from the center of the disk. This assumption is justified if the radius of each hole is small compared with that of the cylinder, but it is impracticable to make it too small as the values of $d$ are thereby reduced. The repeatability with which $d$ could be measured in the experiment was estimated as ±0.002 in. using the same dial gauge. The accuracy with which $\theta$ could be measured in the arrangement employed was $\pm\frac{1}{2}°$.

The gum compound used had the following

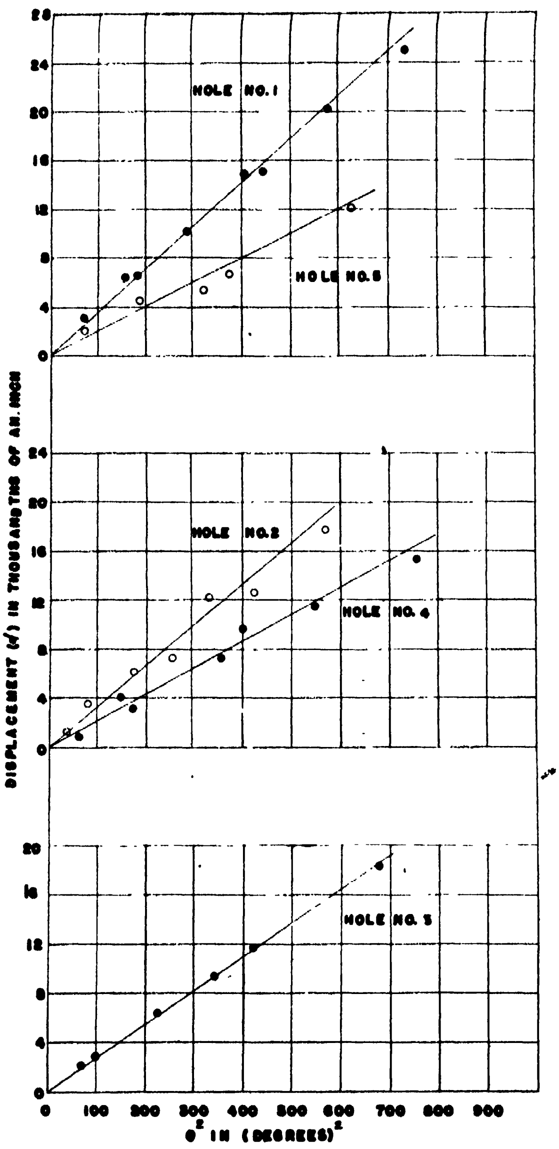


FIG. 5. Relation between displacement and the square of the angle of torsion.

composition in parts by weight:

| | |
|---|---|
| Rubber | 100 |
| Sulphur | 3.0 |
| Captax | 0.8 |
| Stearic acid | 1.0 |
| Agerite | 1.0 |
| Zinc oxide | 5.0 |
| Tuads | 0.08 |

The mix was vulcanized for 30 min. at 134°C. In order to produce a strong bond between the brass plates, a "tie gum" layer, less than $\frac{1}{32}$" thick, was introduced between the main rubber cylinder and each of the brass plates, and a cement marketed as "Reanite, No. 3615" was employed. The composition of the "tie gum" layer was, in parts by weight,

| | |
|---|---|
| Rubber | 100 |
| Stearic acid | 1.0 |
| Sulphur | 3.5 |
| Captax | 1.0 |
| Zinc oxide | 5.0 |
| Thermatomic black | 40 |

The rubber cylinder employed had a radius of 2.12 in. and length of 1 in. The holes in disk $B$ each had a diameter of $\frac{1}{2}$ in. and the distances of their centers from the center of the disk are given in the following table:

| Hole number | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $r$ in inches | 0 | .85 | 1.18 | 1.47 | 1.69 |

Equation (9) suggests that the results of the experiments be plotted as a relation between $d$ and $\theta^2$ on linear graph paper. This should yield a straight line passing through the origin for each hole.

This is done in Figs. 5 for each of the five holes, and it is seen that the linear relation between $d$ and $\theta^2$ of Eq. (9) is borne out within the accuracy of the experiment.

The slope $m$ of the straight line obtained for each hole is the value of $K[2a^2C_2+(a^2-r^2)\times(C_1-2C_2)]$ for that hole, and if this is plotted against $(a^2-r^2)$ a straight line should be obtained, which passes through the origin if $C_2=0$ and otherwise gives an intercept of $2a^2C_2K$ on the $m$-axis. If $m$ is plotted against $(a^2-r^2)$ as in Fig. 6, it is seen that within the accuracy of the experiment the relation (9) between $d/\theta^2$ and $(a^2-r^2)$ is again borne out.

If $m_0$ is the intercept of the straight line in Fig. 6 on the $m$-axis and $b$ is its slope, we have $m_0=2a^2C_2K$ and $b=K(C_1-2C_2)$. Whence

$$\frac{C_1}{C_2}=\frac{2b}{m_0}a^2+2. \tag{11}$$

From measurements on Fig. 6, we obtain

$$m_0=10\times10^{-6}\text{ in. deg.}^{-2},$$

and

$$b=5.65\times10^{-6}\text{ in.}^{-1}\text{ deg.}^{-2}$$

so that

$$\frac{C_1}{C_2}=7.1.$$

Mooney[2] uses the constants $C_1$ and $C_2$ to define three additional constants as follows:

$$G=2(C_1+C_2)$$
$$H=2(C_1-C_2)$$

and

$$\alpha=H/G.$$

He then shows that the constant $G$ is the modulus of rigidity of the material. The constant $\alpha$ is named the coefficient of asymmetry. From the definitions it follows that

$$\alpha=\frac{C_1/C_2-1}{C_1/C_2+1}.$$

Consequently, the experimental results presented here lead to a value of 0.75 for the coefficient of asymmetry of the pure-gum compound studied. This is somewhat larger than the values found for two compounds studied by Mooney. Further work would be required to establish the variation of $\alpha$ with the nature of the compound.

## VI. CONCLUSIONS

The experiments bear out the general conclusion of the theory that in order to produce pure torsion in a rubber cylinder, normal surface tractions must be exerted on the plane ends and that these surface tractions vary with the amount of torsion and the position on the plane end of the cylinder. If the assumption is made that the amount of bulging in each hole is proportional to the force producing it and does not depend on the state of local strain on which the bulging is superposed, then the results of the experiments are in accordance with the assumption of an expression of the form (7) for the stored-energy function of the rubber. It has already been pointed out by Mooney that a stored-energy function of this form describes the linear shearing stress *vs.* amount of shear relation for rubber in simple shear, with a fair measure of accuracy, up to reasonably large deformations. Also, Treloar[4] has recently obtained results in agreement with this for pure, homogeneous strain of a sheet of rubber.

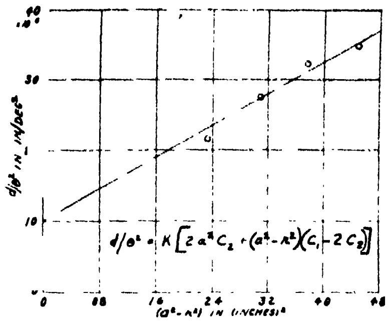


FIG. 6. Relation between slopes of lines in Fig. 5 and positions of holes.

The form (6) for the stored-energy function is that derived from the kinetic theory of rubber-like elasticity. Load *vs.* deformation relations derived from this have been compared with experimental results for simple types of deformation and have yielded fairly good agreement, depending on the type of deformation employed.

It appears then that the form (6) for the stored-energy function provides a first approximation to the stored-energy function for vulcanized rubber and that a second approximation is provided by the form (7). The differences between these forms are more or less accentuated depending on the type of deformation studied.

## VII. ACKNOWLEDGMENTS

This work forms part of a program of fundamental research on rubber undertaken by the Board of the British Rubber Producers' Research Association. The work was carried out at the National Bureau of Standards, Washington, D. C., where the author is at present a guest worker. The author is indebted to W. L. Holt of the National Bureau of Standards for advice and assistance in connection with the mechanical design of the apparatus.

---

[4] L. R. G. Treloar, Trans. Faraday Soc. (pending publication).

# Elastic Losses in Some High Polymers as a Function of Frequency and Temperature*

H. S. SACK, J. MOTZ,** H. L. RAUB, AND R. N. WORK
*Cornell University, Ithaca, New York*

The frequency and temperature dependence of the elastic losses of polyvinyl chlorides (0° to 80°C, 60 to 1400 c.p.s.) and of natural rubber and butadiene-styrene copolymers (GRS) of different relative concentration (−50° to 35°C, 20 to 3500 c.p.s.) was studied by 3 methods, in which the motion of a reed in free or forced vibration is observed. Natural rubber at temperatures above 20°C, and polyvinyl chloride exhibit frequency independent losses, the mechanism of which is not yet clear. The results for rubber below 20° and for GRS can be explained by a relaxation theory. The behavior of the elastic losses also reflects the existence of a second-order transition, below which rotation of chain segments is hampered.

## 1. INTRODUCTION

AMONG the theories of elastic losses in solids, the so-called relaxation theory has lately received considerable attention.[1] According to this theory the reaction of a body to an applied strain is not instantaneous but requires a finite time which is determined by the molecular constitution of the material.[2] The time which is characteristic of the speed with which molecules will move towards their new equilibrium position when a stress is set up is called the relaxation time. This theory is the analogue to Debye's[3] theory of anomalous dispersion of dielectrics, and has been discussed for simple and more complicated models in many publications. It will suffice, here, to recall that according to this theory the elastic losses—defined as energy loss per cycle (often called $Q^{-1}$)—should rise linearly with frequency $\omega$ for low frequencies, and decrease as $1/\omega$ for high frequencies. The shape of the maximum in the intermediate region depends on the number of relaxation processes and, therefore, on the relaxation times that are present. This analogy between dielectric and elastic properties has been emphasized particularly in a recent publication of Kirkwood,[4] in which he applied to elastic phenomena a treatment which he formerly had applied to dielectric behavior where he calculated the distribution of relaxation times on the basis of a diffusion process of the single groups of the long chain molecules in high polymer substances.[5]

So far, little experimental evidence for the existence of this relaxation phenomena has been found, at least in experiments in which the elastic losses under alternating stresses are measured.[6] The aim of the work here presented was to find cases in which the relaxation theory would apply. High polymer molecules seem particularly suited for such a study because these molecules show a high degree of mobility, and their motion seems to be sufficiently slow so that it can be hoped that a relaxation time may be found in a low frequency region. It is also known from dielectric measurements that such relaxation phenomena do exist in these substances.

As a starting point the series of polyvinyl chlorides were chosen because Fuoss[7] observed dielectric relaxation in a region around 1000 c.p.s. However as will be discussed later, no elastic

---

* Presented at the meeting of the Division of High Polymer Physics, American Physical Society, January 30–February 1, 1947. Early results of this investigation were presented at the APS meeting in November, 1943. It was not possible to publish these results earlier except for a short summary (H. S. Sack, J. Motz, and R. N. Work, J. App. Phys. **15**, 396 (1944).) As they are in direct connection with the paper presented at the January 1947 meeting, and in order to avoid duplication in publication, the former work, in its essential lines, is included in the present contribution. Part of the later work was sponsored by the Office of Naval Research.

More details can be found in Cornell theses of J. Motz, R. N. Work and H. L. Raub.

** Now at the Armour Research Foundation, Chicago, Illinois.

[1] R. Simha, J. App. Phys. **13**, 147 (1942); T. Alfrey and P. Doty, J. App. Phys. **16**, 700 (1945); W. Kuhn, Zeits. f. physik. Chemie **B42**, 1 (1939).

[2] The relaxation theory can also be applied in cases where the relaxation is not due to a molecular rearrangement, but rather to heat flow. In this case the relaxation time is a function of the boundary conditions and therefore of the external shape of the body or of the grain structure of the substance. In this article we will refer only to relaxation phenomena which are governed by the motion of the molecules. C. Zener, Metals Tech. [5], **13** (1946). K. Bennewitz and H. Rötger, Physik. Zeits. **37**, 578 (1936); **40**, 416 (1939).

[3] See e.g. P. Debye and H. Sack; Handb. Rad. (1934); Vol.VI, part 1, p. 141.

[4] J. G. Kirkwood, J. Chem. Phys. **14**, 180 (1946).

[5] R. M. Fuoss and J. G. Kirkwood, J. Am. Chem. Soc. **63**, 385 (1941).

[6] The relaxation theory has been tested for nonperiodic motions, as for instance in the case of creep, etc.

[7] R. M. Fuoss, J. Am. Ch. Soc. **63**, 369 (1941).

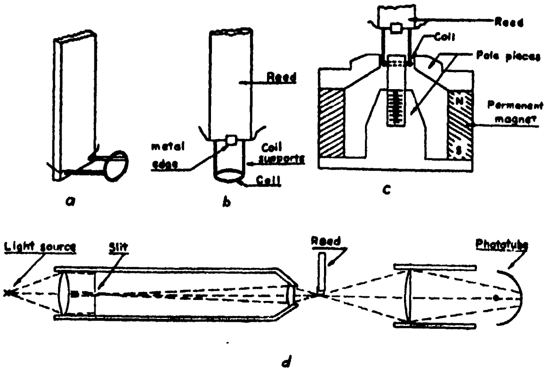


FIG. 1. Diagram of experimental arrangement. (a) and (b) coil attachment; (c) arrangement of magnet; (d) optical system.

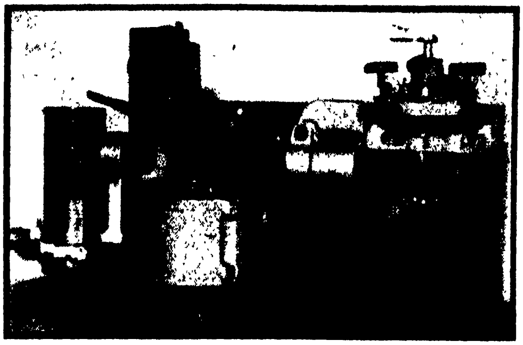


FIG. 2. Photograph of vibrator-unit.

relaxation was observed. More success was obtained with elastomers, in particular with synthetic rubber. The experimental procedure and results will be discussed in the following.

## 2. EXPERIMENTAL PROCEDURE

Three different methods for the determination of the elastic losses were used: (a) The observation of the half-widths of the resonance curve of a reed in forced vibration, (b) the measurement of the damping properties of a free vibrating body, and (c) the measurement of the (elastic) hysteresis loop.

The first and second methods are only applicable if the damping is relatively low. The last method is, in principal, free from this restriction, but the interpretation of the measurements is easier for high damping. The last method has the advantage, that it permits the measurement of the elastic losses over a wide range of frequencies, without need for changing the sample, while for methods (a) and (b) samples of different lengths have to be studied, in order to measure the elastic losses at different frequencies.

In all three cases, the sample, in the form of a reed approximately 2 mm thick and 4–15 mm wide, and of appropriate length, was clamped on one end, while an alternating force could be applied to the other end. In the first experiments this force was applied by an electrostatic method: One side of the reed was covered with a very thin metal foil and opposite the reed, separated by a few mm, was placed an electrode so that an alternating potential could be applied between the reed and this electrode. This gives rise to an alternating force of twice the frequency of the applied potential. (If a d.c. potential is superimposed, it is possible to create an alternating force of the same frequency.) In later measurements a small coil was attached to the end of the reed, and the reed was so mounted that this coil is in the air-gap of a strong speaker magnet. An alternating current set up in the coil then creates an alternating force on the reed.

The motion (vibration) of the end of the reed is observed by an optical method. The end of the reed protrudes into a light beam. If the reed vibrates, this light beam is thus modulated, and in falling on a photo-cell creates an a.c. signal proportional to the amplitude of vibration. This signal is then amplified. Figure 1 shows diagrammatically the optical arrangements and the attachment of the coil to the reed, in (a) for a lateral vibration, and in (b) for a longitudinal vibration. In the later experiments the whole device was made as small and compact as possible in order to avoid outside disturbances. Figure 2 reproduces a photo of the complete assembly mounted on a solid base plate. This picture represents the case of longitudinal vibration. For flexual vibrations a different clamping block is used. The screws on top of the light-bulb holder serve to adjust the light beam so that maximum modulation and, therefore, maximum sensitivity are obtained. For the study of variations with temperature the complete system was put in an automatically controlled constant-temperature box. (This arrangement is, of course, only workable for temperatures not higher than approximately 70°C.)

With the compact optical system as shown in Fig. 2, it was necessary to define the end of the reed more sharply by attaching to it a thin metal edge. The photo-current is first amplified by a preamplifier driven by batteries, and then by a conventional audiofrequency amplifier. The driving potential or current was delivered by an audio oscillator, and kept constant over the whole frequency range. The frequency could be recorded within 1 percent.

In case (a) (resonance curve), the output of the photo-amplifier was read on an a.c. meter giving a direct measure of the amplitude of vibration, which, when plotted as a function of frequency, gives a resonance curve whose half-width is easily determined. This method is precise for low or moderately high damping. For higher damping, the resonance peak becomes very wide and may be distorted by higher harmonics or parasitic reasonances in the system.

In case (b) (free vibration), the reed is deflected from its rest position by a d.c. potential or current. By opening a switch, the current or potential is interrupted and the reed returns to its equilibrium position executing a damped vibration. The output of the photo-amplifier is applied to the deflecting plates of an oscilloscope. At the time, the switch is opened, starting the vibration of the reed, a linear sweep of adjustable speed is applied to the other set of the plates of the oscilloscope. Thus, on the screen appears the picture of the damped vibration, from which the logarithmic decrement can be calculated by measuring successive amplitudes. Figure 3 reproduces two sets of pictures obtained with two materials of different constitution. The four pictures in each set represent different frequencies of vibration as indicated by the time-scale. The set at the left represents a material in which the damping increases very strongly with increasing frequency, whereas in the set at the right the damping is nearly independent of frequency. Again, this method is only usable for low damping.

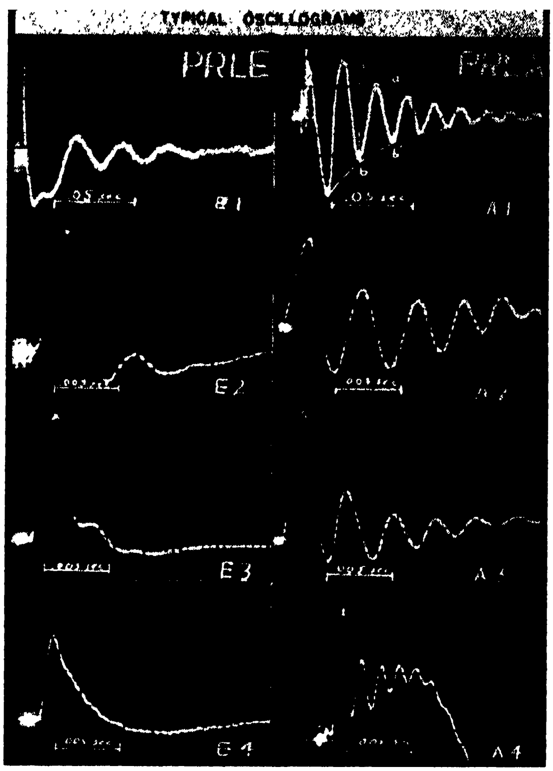


FIG. 3. Examples of damped free vibrations: butadiene-styrene, left 50/50, right 70/30.

In case (c) (hysteresis loop), the procedure is somewhat more complicated. The output of the photo-amplifier is again applied to one set of plates of the oscilloscope. To the other pair of plates a potential is applied which is taken across a resistor in series with the driving coil. Since the driving force is proportional to the current in the coil, this signal is proportional to the force. Thus there will appear on the screen the elastic hysteresis loop of the material (stress-strain diagram). The area of this hysteresis loop is a measure of the damping qualities of the material. In fact, if divided by the square of the amplitude, it is proportional to the energy dissipated per cycle divided by the total energy of vibration. In this method, the two signals must appear on the oscilloscope screen without phase differences caused by the electric system. As phase differences in amplifiers are unavoidable, a phase shifter is added to the system. This phase shifter is calibrated for each frequency by replacing the sample by a metal reed. It is known that for a metal like brass the phase shift between the vibration and the applied force is near 0 or 180°, except for a narrow region around resonance. If with this metal reed the stress-strain curve appearing on the oscilloscope screen is not a straight line, then there exists a phase shift in the electrical system, and it can be compensated for by an adjustment of the phase shifter. The accuracy of this phase adjustment proved to be one of the main difficulties of this method, especially for measurements at frequencies far away from resonance.

In order to have a relative measure of the amplitude, another signal derived from a resistance in series with the coil is passed through the photo-amplifier system. As the current in the coil is kept constant, a variation of this signal as it appears on the oscilloscope screen is a measure of the change in amplification of the photo-amplifier.

The pattern on the oscilloscope screen is photographed, and the area of the hysteresis loop measured with a planimeter. This area is then divided by the square of the maximum amplitude as determined with the help of the calibration line mentioned above.

## 3. INTERPRETATION OF MEASURED QUANTITIES

The quantities determined by the three sets of experiments are: the half-width of the resonance curve $\Delta\omega/\omega_r$ (where $\Delta\omega$ is the difference in frequency for two points where the amplitude of vibration is $\frac{1}{2}$ the amplitude at resonance frequency $\omega_r$); the logarithmic decrement $\delta$; and the energy $W$ dissipated during one cycle, divided by the square of the amplitude $y$, i.e., $W/y^2$. In order to discuss the elastic properties in terms of molecular properties, it is necessary, first, to relate the measured quantities to material constants. In the case of forced vibrations this can be done by introducing a complex modulus of elasticity, e.g., a complex Young's modulus for longitudinal vibrations, $E'+iE''$, which might be frequency dependent. For small damping $(E''/E'\ll 1)$ the same complex modulus can be used in treating the case of free vibrations. In the case of $(E''/E')\ll 1$, the following relations exist $\Delta\omega/\omega_r=\sqrt{3}(E''/E')$; $\delta=\pi(E''/E')$, where $E''$ and $E'$ have the values at the resonance, or free vibration frequency, respectively. $W/y^2$ is proportional to $E''$, with a proportionality constant which depends, in contrast to the two other formulas, on the dimensions of the reed, the boundary conditions etc. (For a free-clamped, thin reed of length $l$ this constant is $\pi^2/8l$.) This proportionality is only valid for frequencies below a certain frequency, which is low for low damping and high, for high damping. For frequencies beyond this limit, deviations from proportionality occur that have to be calculated for each particular case. The same complication arises for the other two methods if the damping is no longer small. In all these cases of nonproportionality it is necessary, in order to calculate the higher approximations to know the exact mechanism which gives rise to the losses, and which determines the frequency dependence of the elastic modulus. These calculations were made under the assumptions of a thin reed, clamped at one end, and acted upon by the external force on the other end, whose motion was observed. The damping mechanism was assumed to be of the viscous type, and the frictional force was written as $K(\partial\epsilon/\partial t)$, where $\epsilon$ is the deformation. In this case $E''=K\omega$. Further calculations were made with putting $E''=\text{const.}$, without specifying the mechanism that would yield this type of loss, since such a mechanism is not yet known. The experimental results were then corrected with the help of the formulas thus derived. Since the assumptions are only approximations to the real case, no data were used where the corrections exceeded 30 percent.

In the hysteresis loop method, the evaluation of $W/y^2$ in absolute units is rather cumbersome. Therefore the width of the resonance curve was measured simultaneously, at least at one temperature. This value was used to calculate the proportionality constants. The results were expressed directly in terms of $\Delta\omega/\omega_r$, and are plotted in these terms in the figures. ($E''(\omega)$ is then the value taken from the graph multiplied by $E'$ at the resonance frequency and at the temperature where resonance is observed, and divided by $\sqrt{3}$.)

## 4. EXPERIMENTAL RESULTS AND DISCUSSION

### (a) Generalities

In all the experiments the amplitude of vibration was so small that the maximum strain was not bigger than 0.1 percent. Under these conditions it proved that all the effects are linear. With the exciting force varying by a factor 5 to 10 in amplitude, no change in the losses was observed. Also the vibration remained sinusoidal (no appearance of harmonics). Furthermore the hysteresis loop has the form of an ellipse within a very good approximation (deviations from the ellipse

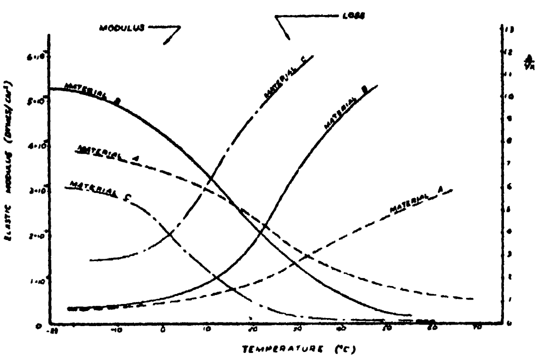


FIG. 4. Young's modulus and elastic losses of polyvinylchloride with 20 percent ($A$), 30 percent ($B$) and 40 percent ($C$) plasticizer.

correspond to an harmonic content of not more than 1 or 2 percent).

The influence of the "tightness" of clamping appears to be negligible for the loss measurement.

The metal foil glued to the reed in the electrostatic method, and the coil and coil supports used in the electrodynamic method, were replaced by similar attachments of different weight or geometric form, and proved to have no influence upon the measured elastic losses.

In most cases measurements were repeated, if possible with reeds of different width or length. Good reproducibility was observed. The error for the losses is approximately ±10 percent.

As will be seen later, very high losses, of an equivalent of approximately 100 for $\Delta\omega/\omega$, were observed under certain conditions. Such high values have to be interpreted with caution, as they mean that all the energy supplied by the outside force is dissipated in a very small region near the point where the force acts, and it is not excluded that the way in which the coil is connected to the reed has an effect; also the formula used for the calculation will probably receive additional correction terms, which cannot be evaluated exactly.

### (b) Polyvinyl Chlorides

Three samples were measured.[8] Material $A$ contained 80 percent polyvinyl chloride and 20 percent tricresyl phosphate, $B$, 70 and 30 percent, and $C$, 60 and 40 percent. All measurements were made by the first method (resonance curve). In Fig. 4 the results for Young's modulus and the half-widths of the resonance curve are indicated as a function of temperature. Young's modulus was calculated from the measured resonance frequency assuming a free-clamped, thin reed (or a clamp-clamped reed) in flexural vibration. By choosing different lengths of the reed, and by vibrating it as free-clamped or clamp-clamped, a series of resonance frequencies are obtained in the range of approximately 100 to 1400 c.p.s. The values obtained at these different frequencies agree within the experimental error with the curves drawn in Fig. 4. We can, therefore, conclude that in this frequency range, no dispersion occurs. This means that in our experiments we do not find a relaxation phenomena as it would have been suggested by the dielectric measurements made by Fuoss on the same substances in the same frequency and temperature range. This does of course not necessarily mean that such relaxation phenomena are absent, but if they are present, they are masked by other processes which are of greater importance and yield frequency independent losses.

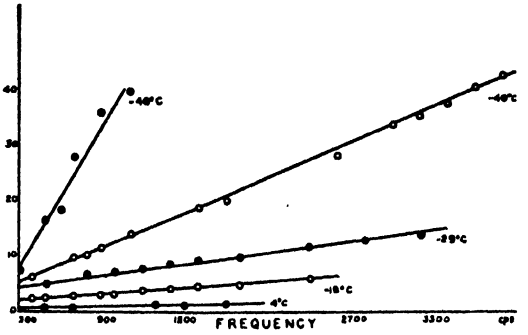


FIG. 5. Losses of natural rubber.

From Fig. 4 it can be seen that Young's modulus decreases with increasing temperature, while the losses increase. This can be explained by the "loosening" of the structure, and the increased tendency to flow. It should also be mentioned that static measurements of Young's modulus, at room temperature and with small stresses, were made, which yielded from 4 to 10 times higher values than those obtained by the vibration method. This seems to indicate that a dispersion zone must be present below the frequency range here studied.

---

[8] We are obliged to Dr. Fuoss, at that time at General Electric, for providing us with these samples.

### (c) Natural Rubber

The sample studied was natural gum stock with 3 percent sulphur. For temperatures between 5 and 55°C measurements were made with all 3 methods, yielding identical resuts; for lower temperature, down to $-46$°C, only the hysteresis-loop method proved workable. The frequency range was 20 to 900 c.p.s. for the first two methods, and 300 to between 1500 and 3500 (depending on the temperature) for the third method. Figure 5 reproduces the results for temperatures below 4°C, and shows that the losses rise linearly with frequency (but the line does not go through the zero). Measurements at temperatures above 10°C are not included in the diagrams. They yield for 20 to 1300 c.p.s., a constant value of 0.2 for 22°C, and slightly lower for higher temperature. (A slight rise of approximately 10 percent would also be compatible with the measurements.)

Where a comparison can be made, our results are in good agreement with those of Stambaugh,[9] and other earlier investigators. The frequency independent losses at higher temperatures can be represented, in a formal way, by a constant $E''$. However, as mentioned, no simple mechanism is known that would yield this result. Some speculations concerning this mechanism were given earlier,[10] but no new evidence in favor of or against it has been obtained. The losses at lower temperatures can be represented by the superposition of a frequency independent part and a viscous term, $E''=K\omega$, where $K$ could be called a coefficient of viscosity. The same frequency dependence would result from a relaxation phenomenon with measurements made below the relaxation time (or zone of relaxation times). In fact, for low frequency, the relaxation theory gives a molecular picture of viscous flow. Figure 6 shows how this coefficient of viscosity increases with decreasing temperature.

Resonance frequencies could be observed above $-18$°C. These values indicate that Young's modulus is rising slightly with decreasing temperature (20 percent increase from 22°C to $-18$°C).

### (d) Butadiene-Styrene Polymers

The samples[11] were unloaded and contained 1 percent sulphur, 0.5 percent BLE and 0.3 percent tetramethylthiuramonosulfide. The content of butadiene to styrene was 70/30; 65/35; 60/40; 55/45; 50/50. For the first 4 samples, at 22°C, all 3 methods were used, and gave identical results. The last sample at all temperatures, and the others below 22° were only investigated by the hysteresis-loop method. As an example Fig. 7 reproduces a set of curves observed for the 60/40 sample. For higher temperatures than those given in the figure, the losses are lower, and the bending is less pronounced. In Fig. 6 the slope of the linear part (low frequencies) is given as a function of temperature for this sample, and for 70/30 and 50/50. It appears that the temperature variation of this slope has the same shape, but is shifted to higher temperatures wih higher styrene content.

The characteristic feature in Fig. 7 as compared to the figure for natural rubber (Fig. 5) is the

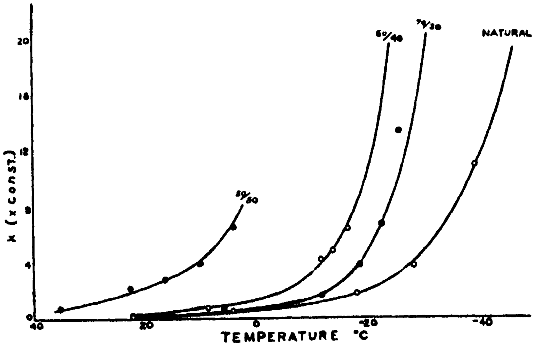


Fig. 6. Temperature dependence of "viscosity coefficient" (slope of loss-frequency curve).

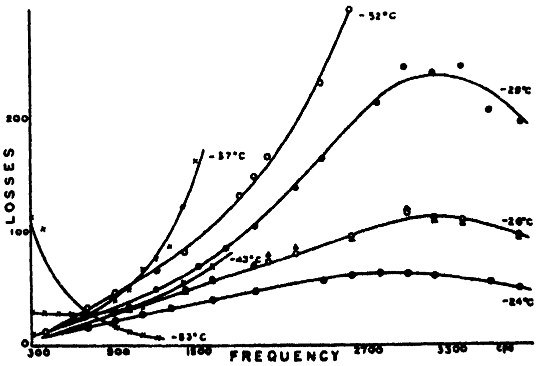


Fig. 7. Losses for 60/40 butadiene-styrene copolymer.

---

[9] R. B. Stambaugh, Ind. Eng. Chem. 34, 1358 (1942).

[10] H. S. Sack, J. Motz, and R. N. Work, J. App. Phys. 15, 396 (1944).

[11] These samples were kindly given to us by the U. S. Rubber Company.

appearance of a peak in the curve. Again, the other samples show a similar behavior, except that the temperature at which the bending of the curve occurs, in the frequency range here studied, is higher for higher styrene content. For the 70/30 sample it does not yet occur above −26°C, while for 50/50 the bending is already pronounced at 16°C. This appearance of a peak can be interpreted as the result of a relaxation phenomenon, though the curves do not correspond exactly to those calculated with a single relaxation time.

Finally it is of interest to note that while at a constant frequency the losses increase first slowly and then more and more rapidly with decreasing temperature, this trend is reversed at very low temperatures (approx. −40°C for 60/40), and the losses decrease again. This fact stands out particularly well in looking at the change of the resonance curve. In lowering the temperature the resonance becomes wider, then is no longer observable, reappears again and becomes narrow. This characteristic temperature is higher for higher styrene content; for 70/30 it could not be observed above −50°C, the lower limit of our temperature range. This characteristic temperature is approximately the second-order transformation temperature of these substances.[12] As it is assumed that below the transformation point the segments of the chain-molecules "freeze in," and are no longer able to orient as freely as above this temperature, we may conclude that the relaxation phenomena observed are linked to this segment mobility. In agreement with this is the observation that the loss-frequency curve below this temperature has a different character from that above, and does not show the characteristic peak, at least not in our frequency range.

---

[12] R. F. Boyer and R. S. Spencer, *Advances in Colloid Sciences* (Interscience Publishers, New York, 1946), Vol. II, p. 1.

---

# Electrical Conductivity of GR-S and Natural Rubber Stocks Loaded with Shawinigan and R-40 Blacks*, **

P. E. Wack, R. L. Anthony, and E. Guth
*University of Notre Dame, Notre Dame, Indiana*

Measurements have been made of the direct current conductivity of rubbers loaded with carbon black. Shawinigan and Continental R-40 blacks compounded in natural rubber and GR-S were studied, and the resistivities were determined as functions of time, temperature, concentration and elongation. Resistance decreased with time, at first very rapidly, then more slowly, approaching an equilibrium value. This behavior seems to be independent of the type of black used. Temperature coefficients of resistance (at 50°C) were positive for Shawinigan stocks, negative for samples containing R-40, and tended to increase with increasing concentration of black and with increasing extension of the sample. At low concentrations, R-40 gave higher conductance in GR-S than did Shawinigan. At high loadings, Shawinigan gave the higher conductance in both GR-S and natural rubber. Resistance increased with elongation for all stocks containing R-40. The Shawinigan GR-S samples showed an inversion; for small elongations the resistance increased on stretching, but for higher extensions the resistance decreased. The results are interpreted on the assumption that the carbon black tends to form chains in the rubber matrix.

## I. INTRODUCTION

Carbon black has been used as a reinforcing agent in rubber-like materials for many years. In recent years special grades of carbon black have been developed which serve the purpose of not only reinforcing these materials but also of imparting to them some electrical conductivity. However, our knowledge of the exact mechanisms or processes, which are responsible for this reinforcing action and greatly increased electrical conductivity, is still far from complete. Both effects are closely related, and increased knowledge of one will aid our understanding of the other.

The reinforcing action of fillers has been dis-

---

* This work was carried out under contract to the Office of Naval Research as a part of Contract N6-ori-83, Task Order 1.

** Presented at the American Physical Society Meeting, Division of High Polymer Physics, held at Columbia University on Jan. 30–31 and Feb. 1, 1947.

cussed in a series of theoretical papers by Rehner,[1] Smallwood,[2] and Guth.[3] A general discussion of the effects of various types of carbon black in GR-S has recently been given by Parkinson.[4] The apparent tendency of the carbon black particles to form chains in the rubber, and possibly even a network structure at high carbon black concentrations has been pointed out by the aforementioned authors and others. Parkinson, in particular, has noted that, in some instances, there is strong evidence that the forces between carbon particles are greater than the forces between the carbon particles and the rubber. The possibility that in some cases the chain of carbon black particles may serve as the major reinforcing agent rather than the individual carbon black particles, was suggested by Guth.[3]

Past studies of the electrical conductivity produced in rubbers by carbon blacks have been confined chiefly to investigations of the effects resulting from varying the type of black, the particle size of the black, and the concentration of the black in the rubber. Cohan and Mackey[5] and Cohan and Steinberg[6] have shown that the specific resistance of a rubber compound at 20% volume concentration decreases with increasing purity of the black, decreases with increasing graphitic structure, and decreases exponentially with decreasing particle size. The progressive decrease in resistance with decreasing particle size had also been noted by Amon and Brown[7] in a study of the electrical resistance of carbon black inks. Hall, Buckley, and Griffith[8] have studied the dependence of the electrical resistance upon the concentration and structure of the black, using Shawinigan black in natural rubber and in GR-S. Bulgin,[9] in one of the most comprehensive papers published to date on this subject, discussed the dependence of the electrical conductivity in conducting rubbers upon the concentration of the black in rubber, its dependence upon the temperature, and its dependence upon the extension. Bulgin[9] employed a large variety of blacks and studied the behavior of these blacks in GR-S, GR-I, and in natural rubber.

In the work reported here, the conductivity has been studied much more completely than in any previous work as a function of time, temperature, extension, and concentration of the black in the rubber. The behavior of two types of conducting blacks, Shawinigan and Continental R-40, in both GR-S and in natural rubber, has been studied. The conductivity of the blacks themselves was also investigated as a function of temperature, under a pressure of approximately 2000 lb. per square inch.

This more complete study was carried out in the expectation that the results might indicate the state of aggregation of the black particles in the rubber and thus lead to conclusions about the relative strength of the carbon black-carbon black and the carbon black-rubber interaction. Since rubber is an insulator, appreciable conduction in a rubber stock loaded with conducting carbon black spheres can take place only if these spheres form a chain. Such chains in turn may form a lattice or network. Our results can be understood on the basis of a simple picture of the formation and breaking up of chains of carbon black particles. Carbon black particles are colloidal in size. To a first approximation, we may neglect the forces between them. Then we can consider the chain formation as a purely

TABLE I.

| GR-S stocks | | Natural rubber stocks | |
|---|---|---|---|
| | Parts by weight | | Parts by weight |
| GR-S | 100.0 g | Smoked sheet | 100.00 g |
| Zinc oxide | 5.0 g | Zinc oxide | 7.85 g |
| Pine tar | 5.0 g | Pine tar | 3.00 g |
| Stearic acid | 2.0 g | Stearic acid | 3.30 g |
| Sulfur | 2.0 g | Sulfur | 2.81 g |
| Captax | 1.5 g | Captax | 0.743 g |
| (Shawinigan or R-40) | variable | (Shawinigan or R-40) | variable |
| Cured for 70 minutes at 40 lbs./in.$^2$ steam pressure. | | Cured for 90 minutes at 35 lbs./in.$^2$ steam pressure. | |

---

[1] J. Rehner, J. App. Phys. **14**, 638 (1943); Rubber Chem. Tech. **17**, 865 (1944).

[2] H. M. Smallwood, J. App. Phys. **15**, 758 (1944); Rubber Chem. Tech. **18**, 292 (1945).

[3] E. Guth, J. App. Phys. **16**, 20 (1945); Rubber Chem. Tech. **18**, 596 (1945).

[4] D. Parkinson, Trans. Inst. Rubber Ind. **21**, 7 (1945); Rubber Chem. Tech. **19**, 100 (1946).

[5] L. H. Cohan, and J. F. Mackey, Ind. Eng. Chem. **35**, 806 (1943); Rubber Chem. Tech. **16**, 918 (1943).

[6] L. H. Cohan, and M. Steinberg, Ind. Eng. Chem. **36**, 7 (1944).

[7] F. H. Amon, and O. J. Brown, American Ink Maker, November, 1941.

[8] R. H. Hall, B. P. Buckley, and T. R. Griffith, Can. Chem. and Process Industries **29**, 587 (1945).

[9] D. Bulgin, Trans. Inst. Rubber Ind. **21**, 188 (1945); Rubber Chem. Tech. **19**, 667 (1946).

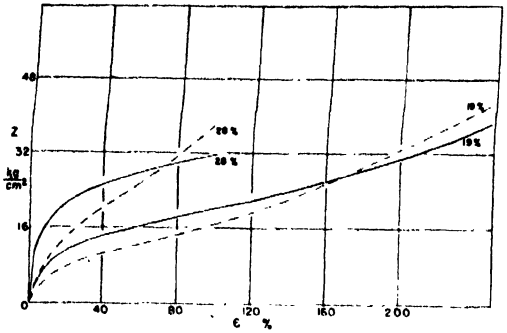


FIG. 1. Fast stress-strain curves at 30°C for GR-S stocks. Shawinigan stocks are denoted by dashed lines and R-40 stocks by full lines.

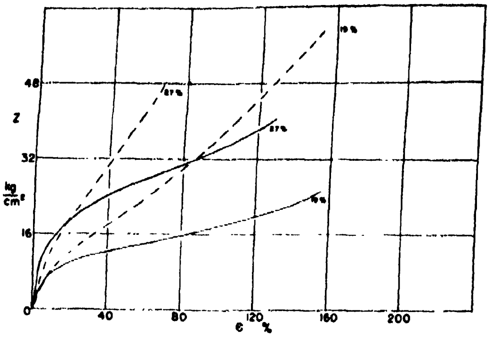


FIG. 2. Fast stress-strain curves at 30°C for natural rubber. R-40 stocks are denoted by full lines and Shawinigan stocks by dashed lines.

kinetic phenomenon regulated only by chance. The probability of formation of a chain becomes appreciable only when the loading is sufficiently high. Changes in temperature and elongation cause a modification of structure of the carbon black chains. This will also change with time. It should be pointed out that the carbon black particles are connected to each other by the flexible long rubber molecules. In the unstretched state, the rubber molecules are curled up, in accordance with the kinetic theory of rubber elasticity. The flexibility of the rubber molecules will permit almost free translational kinetic (Brownian) motion of the carbon black particles, up to the limit given by the length of the rubber molecules. According to this picture, stretching of a loaded stock will limit the region in which the carbon black particles can move, and at sufficiently large extensions, the positions of the carbon black particles are more or less rigidly fixed.

## II. EXPERIMENTAL PROCEDURE

All stocks were mixed on a small laboratory size mixing mill and cured in a small platen press. The following base formulae were used for all stocks: Four different loadings of Shawinigan and of R-40 were employed in the GR-S and natural rubber stocks. These were 136.9 g, 80 g, 50 g, and 36.9 g. These loadings gave volume concentrations of 40 percent, 28 percent, 19 percent, and 15 percent, respectively, for the GR-S stocks, and 39 percent, 27 percent, 19 percent, and 15 percent, respectively, for the natural rubber stocks, the slight difference at the two highest loadings being caused by the difference in densities of GR-S and natural rubber. The volume concentration is defined here as the ratio of the true volume of the carbon black to the total volume of the cured rubber. The carbon black volume was calculated from its weight, employing the standard density value of 1.80 g/cm³. The densities (and volume) of the cured rubber stocks were determined by hydrostatic weighing at room temperature. It is of interest to note that the volume of cured loaded stock is equal to the sum of the volumes of the cured gum stock and of the carbon black, to within 1 percent or better, when the carbon black volume is calculated from the density value of 1.80 g/cm³. This was checked

TABLE II. GR-S stocks.

| Black | Volume concentration (%) | Tensile strength (lb/in.²) | Elongation at break (%) | Shore hardness (30 sec value) |
|---|---|---|---|---|
| Shawinigan | 40 | 1270 | 50 | 97 |
| Shawinigan | 28 | 2070 | 300 | 74 |
| Shawinigan | 19 | 1430 | 750 | 57 |
| Shawinigan | 15 | 1080 | 875 | 46 |
| R-40 | 40 | 1640 | 150 | 97 |
| R-40 | 28 | 2880 | 675 | 74 |
| R-40 | 19 | 2820 | 775 | 61 |
| R-40 | 15 | 3060 | 925 | 50 |

TABLE III. Natural rubber stocks.

| Black | Volume concentration (%) | Tensile strength (lb/in.²) | Elongation at break (%) | Shore hardness |
|---|---|---|---|---|
| Shawinigan | 39 | 1220 | 50 | 99 |
| Shawinigan | 27 | 1630 | 150 | 81 |
| Shawinigan | 19 | 2060 | 575 | 67 |
| Shawinigan | 15 | 2640 | 500 | 66 |
| R-40 | 39 | 1340 | 150 | 95 |
| R-40 | 27 | 2060 | 425 | 78 |
| R-40 | 19 | 2500 | 425 | 70 |
| R-40 | 15 | 2140 | 550 | 60 |

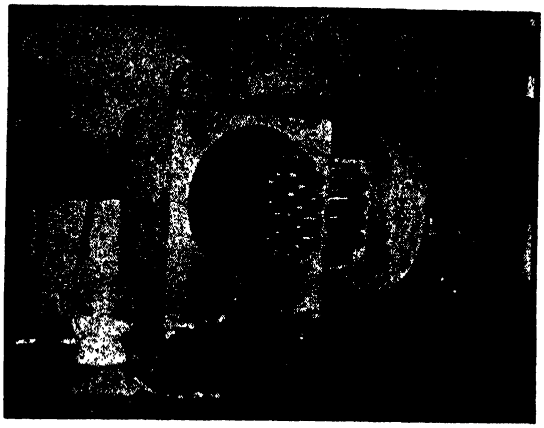


FIG. 3 View of oven with door opened to show mounting rack.

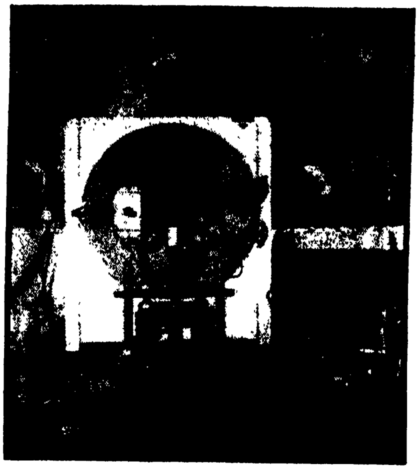


FIG. 4. View of oven with door closed. The apparatus used for measuring resistivities of the blacks is seen in front of the oven.

by hydrostatic weighing of gum stocks and loaded stocks.

The tensile strength, elongation at break and Shore hardness are given in Table II below for the eight GR-S stocks, and in Table III for the eight natural rubber stocks.

Figures 1 and 2 show stress-strain curves at 30°C for the 19 percent and 28 percent concentrations in GR-S, and for the 19 percent and 27 percent concentrations in natural rubber. It is seen that for each loading the R-40 gives higher rigidity in both natural rubber and GR-S at low extensions than does the Shawinigan. At high extensions, the situation is reversed. An explanation for the higher rigidity of Shawinigan stocks at higher extensions will be suggested in connection with the interpretation of resistivity-extension curves (cf. Fig. 14). The data shown in Figs. 1 and 2 were taken with the fast stress-strain machine described in an earlier paper.[10]

The rubber was cured in the form of standard tensile sheets 6 inches square by $\frac{1}{16}$ inch thick. After an aging period of 24 hours or more, the samples for investigation were die-cut from these sheets. These samples were in the shape of square headed dumb-bells, the straight center sections having a width of 0.25 inches and a length of 2 inches. The ends of the samples were placed in flat aluminum clamps which covered only the enlarged end portions of the samples. The clamps served simultaneously as a means of stretching the rubber and also as electrical contacts to the sample. Resistivities of (stretched) samples were measured in the direction of the stretch. The coating of the sample ends with Aquadag before clamping made no appreciable difference in the over-all resistance measurements. After clamping, the samples were mounted in an oven by means of a rack capable of holding eighteen samples, each sample being individually mounted in a horizontal Pyrex glass tube. One end of each sample was held fixed at one end of the glass tube, the other end being attached to a copper rod which protruded through the door of the oven. The elongation of each sample could be adjusted and measured by means of these rods. The time required, after a temperature change, for equalization of temperature throughout the oven, including the interior of the samples themselves, was checked by means of a differential thermocouple. This waiting time was approximately 15 minutes after reaching the desired temperature, as read on a mercury thermometer centrally located in the oven. Good air circulation was maintained by means of a small fan mounted in the base of the oven. Resistances were measured with the aid of a Leeds and Northrup Type H.S. galvanometer and a 2-volt storage battery. For low resistance samples a Leeds and Northrup Testing Set, Model 5430, was also used. Specific resistances were then calculated for each sample from the known sample dimensions.

---

[10] S. L. Dart, R. L. Anthony, and P. E. Wack, Rev. Sci. Inst. 17, 106 (1946).

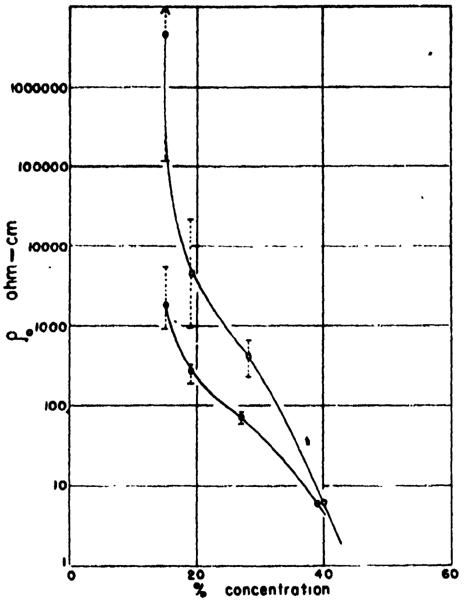


FIG. 5. Resistivity *versus* volume concentration of Shawinigan black at 70°C. Open circles denote GR-S stocks, and black circles denote natural rubber stocks.

The measurements on the carbon black were made in the oven by inserting the powder in a small Bakelite tube and maintaining a pressure of approximately 2000 lbs./in.$^2$ on the black by means of a piston and tension spring device. This pressure, along with considerable tamping down of the black while inserting it in the cylinder, gave apparent densities ranging from 0.36 g/cm$^3$ to 0.39 g/cm$^3$.

Figure 3 shows the oven with its door opened, and the rack in which the samples are placed. Figure 4 is a view of the oven with its door closed and shows the copper rods projecting through the door. The device employed in measuring the temperature coefficients of resistivity for the carbon blacks is seen in front of the oven.

Reversal of the direction of current flow during the measurements, as well as several checks employing 60 cycle current, gave no evidence of any important polarization effects, either in the rubber or in the carbon black.

## III. RESULTS

### A. Dependence of Resistivity Upon Carbon Black Concentration and Upon Time

Initial measurements of the resistivities of several samples showed that the resistivity for each sample was a function of the time which had elapsed after the sample had been cut out and mounted in the clamps in an unstretched condition. In all cases the resistance decreased with time, at first quite rapidly, and then more slowly, and appeared to be approaching a stable value. This behavior is easily explained using our picture of the rubber-black system. As the samples were molded, cured, and cut, strains were set up in their interior. These strains caused a breakage in the carbon black chains, thus leading to an increase in the resistance. In the course of time kinetic (Brownian) motion of the carbon black particles will lead to a re-forming of the chain structure and thus to a decrease in the resistivity. Any small temporary distortion of the sample, for example a small extension followed by an immediate retraction to zero extension, produced a relatively large increase in the resistance, which then decayed in the same manner with time again to a semi-equilibrium value. This behavior had been noted much earlier by Evans[11] and has since been pointed out by Bulgin and others. The approach to a steady state value of the resistance can be greatly accelerated by "annealing" the samples at elevated temperatures. This effect would seem to indicate that any sudden change in

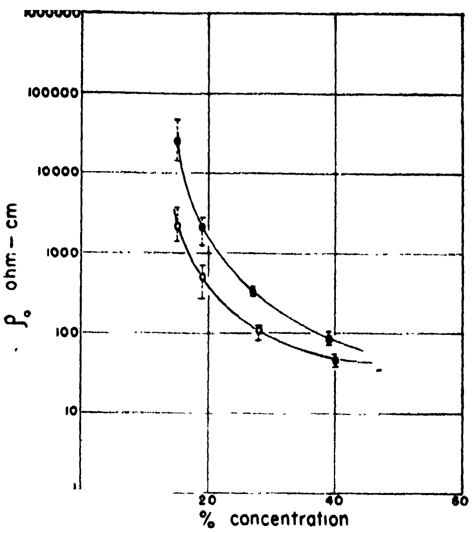


FIG. 6. Resistivity *versus* volume concentration of R-40 black at 70°C. Open circles denote GR-S stocks, and black circles denote natural rubber stocks.

---

[11] R. D. Evans, Paper presented at the meeting of the Ohio Academy of Science in 1930 and kindly made available to us by W. W. Vogt.

strain disrupts existing carbon networks, which subsequently tend to reform themselves. In the work described here all samples were annealed, in an unstretched condition, for a period of 10 hours at 70°C after they had been mounted in the oven. The value of the resistivity attained at the end of this initial annealing period is denoted by $\rho_0$ in this work. (It should be noted here that this value of resistivity is not really constant but rather continues to decrease slowly with time. However, the important consideration is that the rate of change of resistance with time must be so small that it cannot appreciably alter the results of some additional experiment.) Preliminary measurements also indicated wide variations between the resistivity values for samples which were supposed to be identical. Resistivities for samples cut from the same tensile sheets were found to vary, in some cases, by as much as a factor of ten. This was particularly true for the Shawinigan black– GR-S stocks at low carbon concentrations. The Shawinigan black—natural rubber stocks were somewhat more uniform. However, the continental R-40 stocks were much superior in this respect. At high concentrations the variations in resistivity values were much smaller for all stocks.

The dependence of $\rho_0$ upon the volume concentration of the carbon black in the rubber is shown in Figs. 5 and 6. Figure 5 shows the effect on resistivity of four concentrations of Shawinigan in both GR-S and natural rubber, and Fig. 6 shows the behavior of R-40 in these two rubbers. In these two figures each point is an average value of a large number of separate determinations on individual samples. The vertical line drawn through each point shows the range of $\rho_0$ values for each stock. Comparison of Figs. 5 and 6 shows that at low loadings R-40 gives better conductivity (lower resistivity) in GR-S than does Shawinigan. At the high loadings, Shawinigan gives better conductivity in both GR-S and natural rubbers. The R-40 natural rubber stocks have higher resistivities than the R-40 GR-S stocks. In the case of the Shawinigan black, however, the situation is reversed and here the GR-S stocks have the higher resistivities. The very rapid decrease of resistivity with increasing concentration of Shawinigan black at low concentrations, may explain in part the wide variations found in the resistivity values for identical stocks in this region. Slight variations in concentration of the Shawinigan, from sample to sample, might thus be expected to produce large changes in resistivity in the region of low loadings. Large variations in the resistivities of supposedly identical rubber stocks loaded with Shawinigan black were also found by Hall, Buckley, and Griffith.[8] They suggest that the effect may be due to variations in the degree to which the structure of the Shawinigan black has been broken down during the milling and curing processes.

To minimize the effects of variations in the value of $\rho_0$ and thus make possible better comparisons between different samples, when studying the dependence of resistivity upon extension, temperature and time, the relative resistivity, $\rho/\rho_0$, is employed in this work rather than the actual resistivity, $\rho$. Figure 7 shows the decrease in the relative resistivity with time during the annealing period for 19 percent volume concentrations of R-40 and Shawinigan in natural rubber. Figure 8 shows the decrease in this ratio

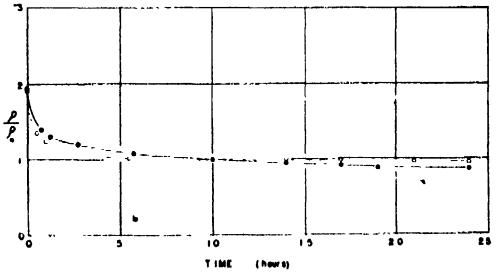


FIG. 7. Annealing curves at 70°C for 19 percent volume concentrations of R-40 (open circles) and Shawinigan (black circles) in natural rubber.

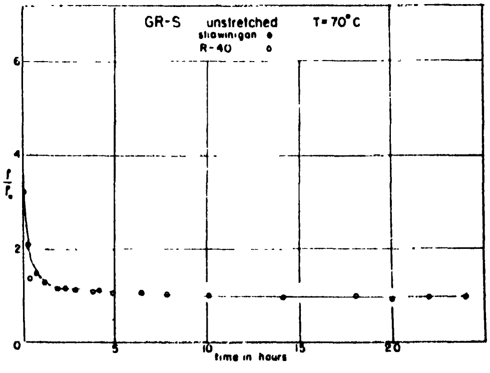


FIG. 8. Annealing curves at 70°C for 19 percent volume concentrations of R-40 (open circles) and Shawinigan (black circles) in GR-S.

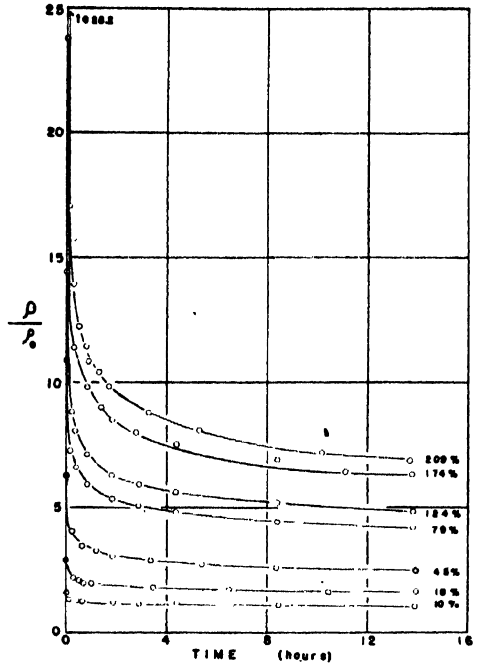


FIG. 9. Decay of resistivity with time at various extensions for a 19 percent volume concentration of R-40 in GR-S. Temperature = 70°C.

with time, during the annealing period, for the same concentrations of these blacks in GR-S. In general, this relative decrease with time was found to be larger for the GR-S stocks than for the natural rubber stocks, and it became progressively smaller for both stocks as the loadings were increased.

## B. Dependence of Resistivity of Loaded Stocks upon Extension

At the end of the 10-hour annealing period, the samples were stretched to various elongations, and the temperature was maintained at 70°C for an additional 14 hours. During this period the decay of resistivity with time was again observed. Figure 9 shows the decrease in the relative resistivity with time, at various extensions, for a GR-S stock containing a 19 percent volume concentration of R-40. These curves have been cross-plotted at times of 30 seconds and 14 hours to show the dependence of the resistivity upon the extension. This dependence is shown in Fig. 10. Figure 11 shows the decrease in $\rho/\rho_0$ with time, at various extensions, for a 19 percent concentration of R-40 in natural rubber. The

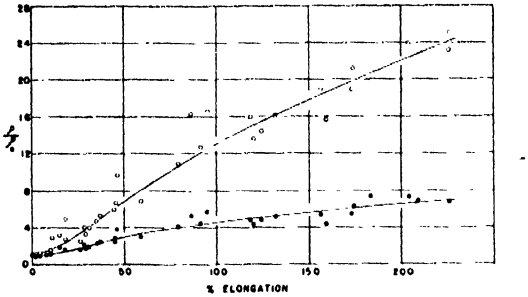


FIG. 10. Cross-plot of Fig. 9 showing the dependence of resistivity upon extension at times of 30 sec (open circles) and 14 hours (black circles).

corresponding dependence of $\rho/\rho_0$ upon the extension, at times of 30 seconds and 14 hours, is shown in Fig. 12. A very similar behavior was observed for the 15 percent and 28 percent concentrations of R-40 in GR-S and also for the 15 percent and 27 percent concentrations of R-40 in natural rubber. It was not possible to obtain data as a function of extension for the 40 percent loadings in GR-S nor for the 39 percent loadings in natural rubber, because of the extreme stiffness of these stocks.

Figure 13 gives the dependence of $\rho/\rho_0$ upon time for various extensions of a GR-S stock containing a 19 percent volume concentration of Shawinigan. A very pronounced inversion in the value of the relative resistivity as a function of

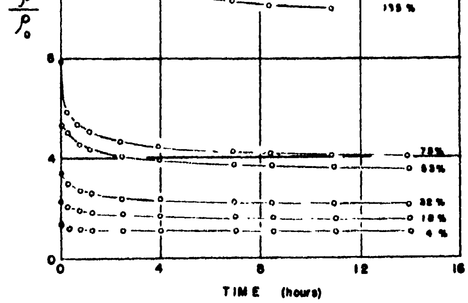


FIG. 11. Decay of resistivity with time at various extensions for a 19 percent volume concentration of R-40 in natural rubber. Temperature = 70°C.

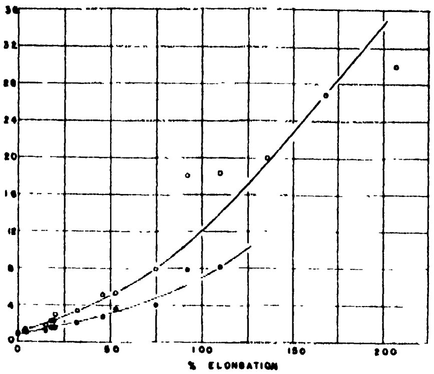


FIG. 12. Cross-plot of Fig. 11 showing the dependence of resistivity upon extension at times of 30 sec (open circles) and 14 hours (black circles).

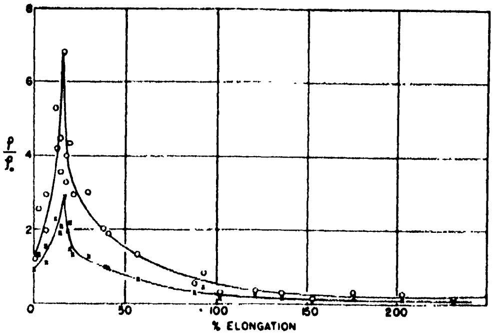


FIG. 14. Cross-plot of Fig. 13 showing the dependence of resistivity upon extension at times of 30 sec (open circles) and 14 hours (crosses).

extension is apparent in the range of low extensions. These curves have been cross-plotted at times of 30 seconds and 14 hours to show the dependence of $\rho/\rho_0$ upon the extension. This dependence is shown in Fig. 14. Examination of Figs. 13 and 14 shows that for high extensions the resistivities are almost immediately reduced to values which are very considerably less than the values which existed just before the extension. At extensions in the vicinity of 15 percent to 20 percent the resistivities, immediately after stretching for the first time, tend to become quite large and are extremely sensitive to slight differences in extension. Reproducibility of relative resistivities in this region was poor. The general behavior shown in Figs. 13 and 14 is strongly suggestive of a maximum in the resistivity *versus* extension relationship in the region of low extensions. This might, perhaps, be caused by a general breakdown of the carbon black network structure at these extensions, followed by a forced realignment of the carbon chains in the direction of the extension at higher elongations caused by the orientation of the rubber molecules. A process of this nature would account for the increased conductivity in the direction of the extension and might even be accompanied by a reduction in the conductivity in the direction perpendicular to the stretch. The latter possibility has not yet been checked, since the equipment employed in this work was designed to measure resistivities (and conductivities) in the direction of the extension.

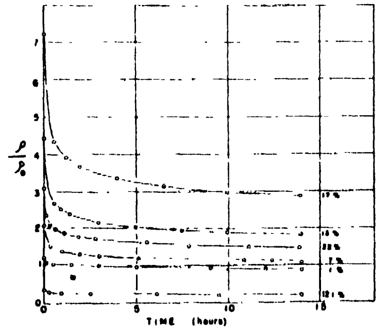


FIG. 13. Decay of resistivity with time at various extensions for a 19 percent volume concentration of Shawinigan in GR-S. Temperature = 70°C.

A probable explanation of the different behavior of Shawinigan and of R-40 at large extensions may be given. At large extensions of the material the rubber molecules most firmly bound to the carbon black particles will be subject to exceptional high extensions and will make abnormally high contributions to the rigidity. Therefore, we would expect carbon black to increase the rigidity of the material more when it is more firmly bound to the rubber molecules. That this is the case with Shawinigan is indicated by the decrease in resistivity below the initial value at high extensions. However the larger size of the Shawinigan particles may also be of importance both for the increase of rigidity and for the decrease of resistance with increasing extension.

An anisotropic conduction of the type suggested above has been shown by Habgood and Waring[12] to occur in extruded stocks containing

---

[12] B. J. Habgood, and J. R. S. Waring, Trans. Inst. Rubber Ind. 17, 51 (1941); Rubber Chem. Tech 15, 146 (1942).

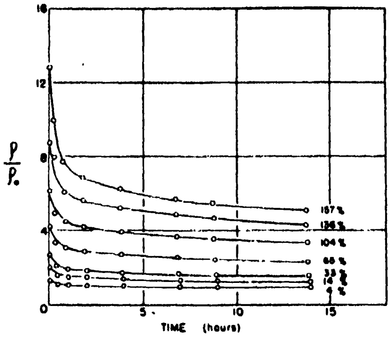


FIG. 15. Decay of resistivity with time at various extensions for a 19 percent volume concentration of Shawinigan in natural rubber. Temperature = 70°C.

Shawinigan black. A similar asymetry in graphite electrodes was also noted by Buerschaper.[13]

The behavior of Shawinigan black in GR-S for the 28 percent loading was similar to that shown in Figs. 13 and 14 for the 19 percent loading. In the case of the 15 percent loading of Shawinigan black in GR-S, the resistivities of the extended samples were too high for measurement with the present equipment. A similar inversion in the resistivity *versus* extension relationship was found for the 27 percent loading of Shawinigan in natural rubber, but was not perceptible for the 15 percent and 19 percent loadings in natural rubber. Figures 15 and 16 show the dependence of relative resistivity upon extension and time for the 19 percent volume concentration of Shawinigan in natural rubber.

## C Dependence of Resistivity of Loaded Stocks upon the Temperature

After the samples had been maintained in an extended condition for a time of 14 hours at 70°C, the temperature was lowered in steps to about 30°C and the corresponding values of resistivity were noted. The temperature was then increased in steps, and readings were taken as a function of increasing temperature. In general the resistance *versus* temperature relationship so obtained was not a linear one, and the two branches of the curve did not coincide. Figures 17, 18, 19, and 20 show the temperature dependence of resistivity at zero extension for all loadings of Shawinigan and R-40 in both natural rubber and GR-S. In general, the agreement between data obtained with decreasing and increasing temperatures becomes better as the loadings are increased. This is particularly true for the Shawinigan stocks. The temperature coefficients of resistivity were determined by constructing tangents to these curves at 50°C, this temperature coefficient being defined as

$$\alpha_{50^\circ\mathrm{C}} = \frac{\Delta\rho}{\Delta T \cdot \rho_{50^\circ\mathrm{C}}}.$$

The dependence of $\alpha_{50^\circ\mathrm{C}}$ upon carbon black concentrations is shown in Fig. 21 for GR-S and in Fig. 22 for natural rubber. Each plotted point is an average of four separate determinations. It is seen that $\alpha_{50^\circ\mathrm{C}}$ is positive for all Shawinigan stocks, and increases with increasing loadings. With one exception, namely, the 40 percent volume concentration in GR-S, the temperature coefficient for the R-40 stocks is negative and also increases with increasing loadings. Since the temperature coefficients at 100 percent concentration should correspond to those of the carbon blacks themselves, the possible trend at higher concentrations is indicated in Figs. 21 and 22 by dotted line extensions of the curves.

Figures 23, 24, 25, and 26 show the dependence of the resistivity-temperature relation upon extension for 19 percent volume concentration of both blacks in GR-S and natural rubber. Average temperature coefficients over the range from 30°C to 70°C have been calculated from these curves as follows. Let $A$ be the average value of $\rho/\rho_0$ at 70°C and $B$ the value of this ratio at 30°C.

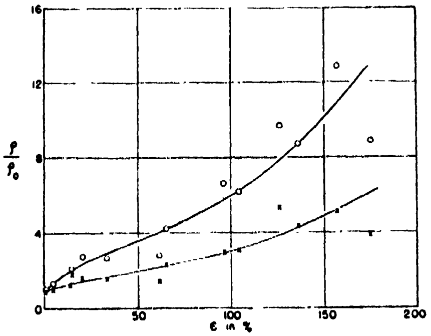


FIG. 16. Cross-plot of Fig. 15 showing the dependence of resistivity upon extension at times of 30 sec (open circles) and 14 hours (crosses).

---

[13] R. A. Buerschaper, J. App. Phys. 15, 452 (1944).

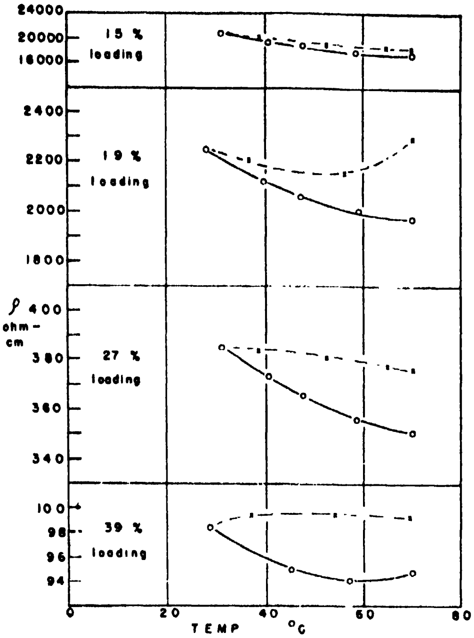


FIG. 17. Resistivity *versus* temperature relationships at zero extension for four loadings of R-40 in natural rubber. Circles indicate decreasing temperature and crosses increasing temperature.

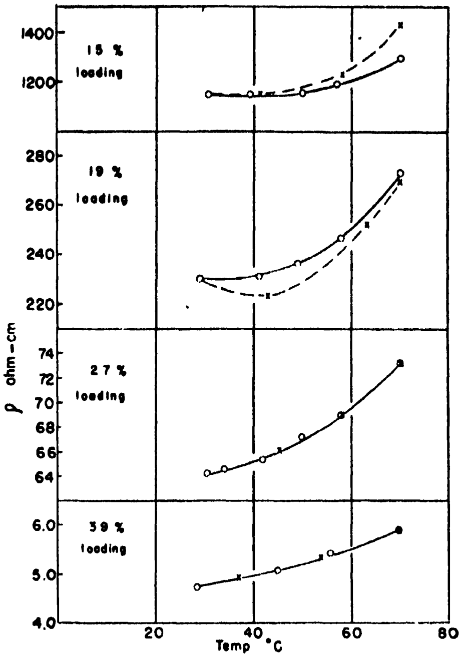


FIG. 19. Resistivity *versus* temperature relationships at zero extension for four loadings of Shawinigan in natural rubber. Circles indicate decreasing temperature and crosses increasing temperature.

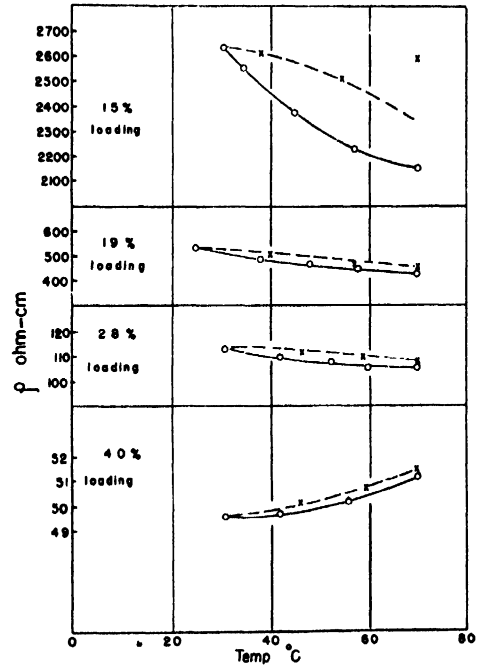


FIG. 18. Resistivity *versus* temperature relationships at zero extension for four loadings of R-40 in GR-S. Circles indicate decreasing temperature and crosses increasing temperature.

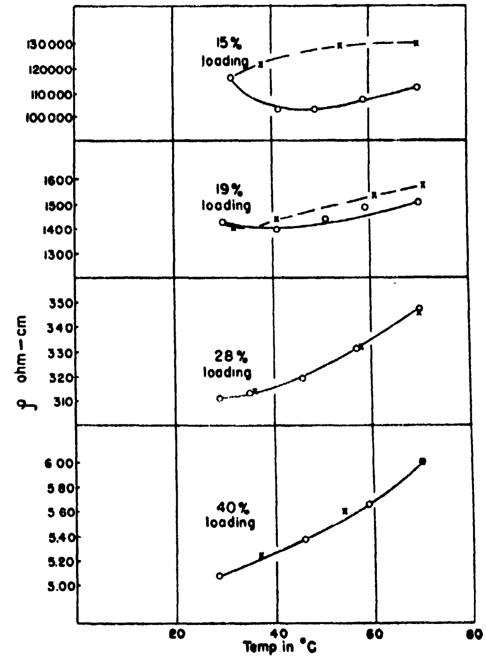


FIG. 20. Resistivity *versus* temperature relationships at zero extension for four loadings of Shawinigan in GR-S. Circles indicate decreasing temperature and crosses increasing temperature.

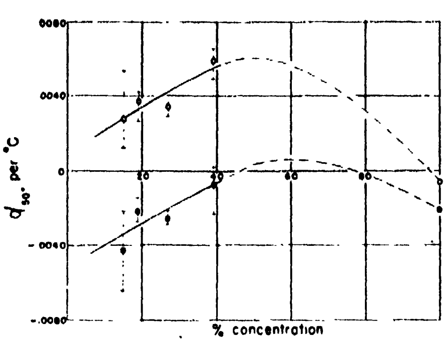


FIG. 21. Dependence of temperature coefficient of resistance upon volume concentration in GR-S. Open circles denote Shawinigan and black circles R-40.

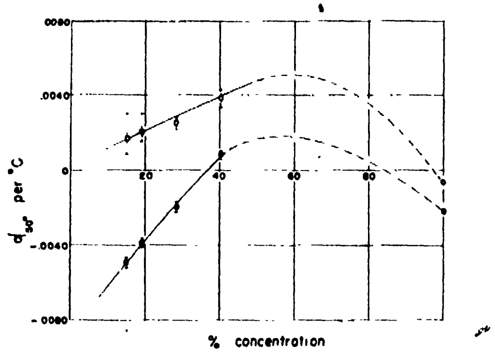


FIG. 22. Dependence of temperature coefficient of resistance upon volume concentration in natural rubber. Open circles denote Shawinigan and black circles R-40.

Then

$$\bar{\alpha}_{50°C} = \frac{A-B}{40\frac{(A+B)}{2}} = \frac{A-B}{20(A+B)}.$$

The dependence of this temperature coefficient upon extension for 19 percent volume concentrations of R-40 and Shawinigan is shown in Fig. 27 for natural rubber and in Fig. 28 for GR-S. This coefficient increases with increasing extension in all cases. It is positive for the Shawinigan stocks and negative for the R-40 stocks. A similar behavior was found for the other loadings.

These results show that the temperature coefficient of the resistivity of the loaded stocks depends strongly upon the concentration and upon the extension. The positive coefficient is probably connected with the breaking and reforming of carbon black chains. Such a kinetic effect will, in general, increase with increasing temperature, in contrast to the resistivity of the black itself, which will decrease with rising temperature. The explanation of the negative coefficient at lower black concentrations needs further consideration. It will also be noted that when a run is made first decreasing the temperature and then increasing it again to the original value the final resistivity is (with one exception: 19 per cent concentration Shawinigan in GR-S) larger than the starting value. This fact and the presence of hysteresis effects in general indicates that any disturbances of the material, whether thermal or mechanical, tends to break up some chains and thus increase the resistivity temporarily. This effect is simply superimposed on the other trends which would be observed if measurements could be made on samples in complete statistical equilibrium. The decrease of the area of hysteresis loops on elongation for R-40 and Shawinigan in GR-S may be explained as due to the fixation of black particles as explained in connection with the effect of elongation on resistivity in general.

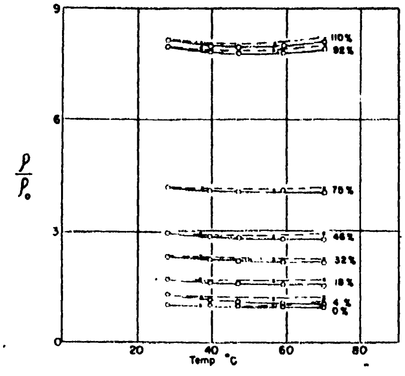


FIG. 23. Relative resistivity *versus* temperature at various extensions for a 19 percent concentration of R-40 in natural rubber.

### D. Dependence of Resistivity of the Carbon Blacks upon Temperature

The temperature dependence of the blacks was measured with the aid of the apparatus described in Section II, a pressure of 2000 lbs./in.$^2$ being maintained on the blacks during the measurements. The dependence of the resistivity upon the temperature is shown in Fig. 29 for both blacks. The temperature coefficients of resistivity at 50°C, as determined from these curves, are $-0.0021$/°C for R-40 and $-0.0006$/°C for Shawinigan.

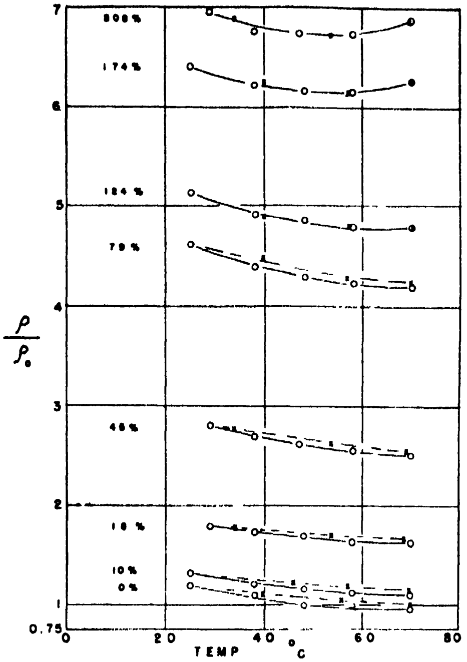


FIG. 24. Relative resistivity *versus* temperature at various extensions for a 19 percent concentration of R-40 in GR-S.

It is interesting to note that even at a pressure of 2000 lbs./in.$^2$ the apparent density of the carbon black ranged from about 0.36 g/cm$^3$ to 0.39 g/cm$^3$, whereas the apparent density of the carbon black in the rubber at the highest loadings was about 0.7 g of carbon black per cubic centimeter of cured rubber. It had originally been expected that pressures of from 1000 to 2000 lbs./in.$^2$ would result in a stable value of resistivity for the carbon black, since a previous study of the decrease in resistance of the carbon black with increasing pressure seemed to indicate a leveling off of resistance values in the region of about 1000 lbs./in.$^2$ On the basis of this expectation, these resistivity values were at first included in the plots of the logarithm of the resistivity *versus* volume concentration shown in Figs. 5 and 6, as points corresponding to a 100 percent volume concentration. This led to an approximate linear relationship for the R-40 black in GR-S and in natural rubber when log resistivity was plotted *versus* log concentration. However, it has since been discovered that the resistivity of the black itself depended strongly upon its apparent density and that even at constant pressure a wide variation in apparent densities can be obtained by tamping down the black while it is being loaded in the cylinder. Since the carbon black volume, calculated from the density value of 1.8 g/cm$^3$, plus the gum rubber volume has been shown to be equal to the volume of the cured loaded rubber, it follows that the "apparent" density of the carbon black corresponding to a 100 percent volume loading should be 1.8 g/cm$^3$. This density value is much greater than any we have been able to produce, and the corresponding resistivity value is therefore unknown. This strong dependence of resistivity upon apparent density of the black has been noted by Benson,

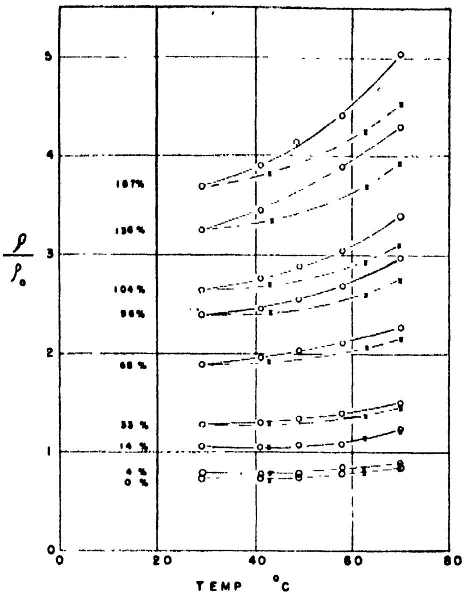


FIG. 25. Relative resistivity *versus* temperature at various extensions for a 19 percent concentration of Shawinigan in natural rubber.

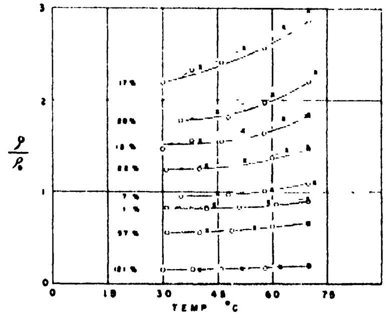


FIG. 26. Relative resistivity *versus* temperature at various extensions for a 19 percent concentration of Shawinigan in GR-S.

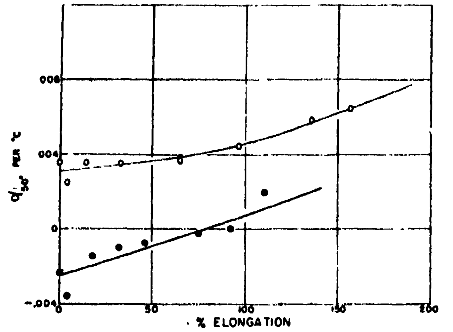


FIG. 27. Dependence of temperature coefficient of resistivity upon extension for 19 percent concentrations of R-40 (black circles) and Shawinigan (open circles) in natural rubber.

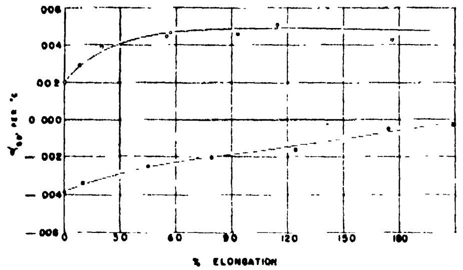


FIG. 28. Dependence of temperature coefficient of resistivity upon extension for 19 percent concentrations of R-40 (black circles) and Shawinigan (open circles) in GR-S.

Gluck, and Kaufmann.[14] They also showed that the density of the black measured at a constant pressure could be increased by as much as a factor of two by various densifying processes, for example by ball-milling.

In contrast to the large variations in the resistivity of the blacks produced by variations in pressure and density, the temperature coefficients of resistivity were found to be nearly independent of pressure and density, provided neither the pressure nor the density was too low. Accordingly these values have been included in Figs. 21 and 22.

## IV. SUMMARY

The behavior of two types of conducting blacks, R-40 and Shawinigan, has been studied in natural rubber and in GR-S. A series of four loadings of each black in each type of rubber was employed. These loadings ranged from about 40 grams of black per 100 grams of raw rubber to 140 grams of black per 100 grams of raw rubber, the corresponding true volume concentrations ranging from 15 percent to 40 percent. The resistivities of these stocks were studied as functions of time, temperature, extension, and loading. The following results were observed:

### A

Fast stress-strain curves show that in both natural rubber and GR-S, the R-40 black gives the better reinforcement at low extensions, but that Shawinigan gives the better reinforcement at the high extensions.

### B

Any small strain produces an increase in resistivity which then decays with time in a nearly exponential fashion. In particular the process of cutting out the samples and mounting them in clamps produces such strains. Annealing the samples at elevated temperatures accelerates the approach to an initial steady-state resistance value. Accordingly, in this work all samples were annealed for 10 hours at 70°C in an unstretched condition. The resistivity value attained at the end of this time is denoted by $\rho_0$.

### C

At the lowest concentrations, R-40 gave better conductivity in GR-S than did Shawinigan, the log resistivity *versus* concentration curve being very steep for the Shawinigan black in this region. At the high loadings Shawinigan gave higher conductance in both GR-S and natural rubber. The R-40 GR-S stocks were better conductors than the R-40 natural rubber stocks. For Shawinigan black, however, the natural rubber stocks were better conductors than the GR-S stocks.

### D

For the R-40 stocks, stretching a sample produced an increase in the resistivity, which then decayed with time while the sample was held at a fixed extension. The Shawinigan stocks behaved differently, however. In the case of the 19 percent and 28 percent concentrations in GR-S and also the 27 percent concentration in natural rubber, an inversion was found in the resistivity *versus* ex-

---

[14] G. Benson, J. Gluck, and C. Kaufmann. In paper presented before the Toronto Meeting of the Electrochemical Society, October 16 to 19, 1946.

tension relationship. The resistivity increased with increasing extensions up to about 20 percent extension and then decreased for higher extensions. For sufficiently high extensions, the resistivities were immediately reduced to values which were very considerably less than the unstretched values. All resistivities were measured along the length of the sample and hence in the direction of the extension.

### E

The dependence of resistivity upon temperature was observed for all stocks over the temperature range from 30°C to 70°C. In general, the average temperature coefficient, calculated at 50°C, was positive for the Shawinigan stocks and negative for the R-40 stocks. It increased with increasing carbon black concentration and also increased with increasing extension of the sample.

### F

Temperature coefficients of resistivity were also measured for the two blacks at 50°C. The blacks were compressed under a pressure of about 2000 lbs./in.$^2$ This pressure was maintained during the measurements, the "apparent" densities ranging from 0.36 to 0.39 g/cm$^3$. The observed temperature coefficients were −0.0021/°C for the R-40 black, and −0.0006/°C for Shawinigan black.

Our results can, in general, be understood on the basis of a simple picture of the formation and the breaking up of chains of carbon black particles. Thermal and mechanical disturbances both lead to a breakage of the chains and thus to an increase in the resistance. In the course of time Brownian motion of the carbon black particles will re-form the chain structure and thus decrease the resistivity as a function of time. This explains the decrease of resistance with time of freshly cured and cut and of stretched samples, the increase of resistivity on stretching in general, the fact that the ascending branch of the resistivity-temperature curve lies above the descending branch and a positive temperature coefficient of resistivity. In the case of Shawinigan the greater increase of rigidity and the decrease of resistivity (below the initial value), both at high extensions, is probably determined by the greater strength of the rubber-carbon black bond and the larger size of the black particles as compared with R-40.

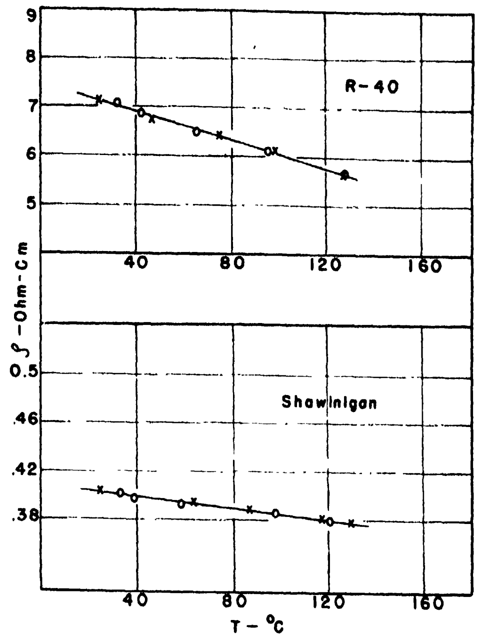


FIG. 29. Dependence of carbon black resistivity upon temperature at a pressure of 2000 lbs./in.$^2$

It is a pleasure to thank Professor Hubert M. James for frequent consultation and many valuable suggestions in connection with this study.

# Elastic Properties of Cork

## II. Stress-Temperature Relationship of Compressed Cork

S. L. Dart and Eugene Guth
*Polymer Physics Laboratory, University of Notre Dame, Notre Dame, Indiana*

(Received October 14, 1947)

In the first paper of this series it was shown that for cork, held at constant compression, the stress as a function of compression, temperature, and time was separable into two factors, one factor representing the load-compression curve and the other describing the decay of stress with time. In the present paper the dependence of the load-compression curve upon temperature is studied. After the sample has relaxed at constant temperature and compression for a time sufficiently long so that the change of stress with time may be neglected, stress-temperature curves were obtained. These curves showed a linear dependence of the stress upon the temperature. From the stress-temperature curves the resolution of the stress into its additive components due to internal energy and entropy was carried out. The physical significance of this resolution is discussed. The experiments were carried out with a refined version of the apparatus described in the first paper. In particular, the apparatus was corrected for rate of thermal expansion.

### INTRODUCTION

In a previous paper[1] an account was given of the stress relaxation of cork as a function of compression, temperature, and time. In particular it was shown that the stress decayed to zero in about 1000 years at room temperature and in about 1000 minutes at 200°C, independent of compression. This independence of the decay time on compression was also checked at 80°C and at 130°C. It was also demonstrated that the compressive stress $Z$ as a function of compression, temperature, and time was separable into two factors,

$$Z(\epsilon, T, t) = F(\epsilon, T)\cdot G(t, T), \tag{1}$$

where the factor $F(\epsilon, T)$ is essentially the load-compression curve of cork and the factor $G(t, T)$ describes the decay of stress with time and temperature. In the previous paper it was proved that $G(t, T)$ has the form

$$G(t, T) = a - b \log t \tag{2}$$

where $a$ and $b$ depend on $T$ only. In particular $a$ and $b$ were thrown into the form

$$a = 1$$
$$b = \beta/(\alpha - T)$$

giving as a final equation

$$Z(\epsilon, T, t) = A(\epsilon, T)[1 - [\beta/(\alpha - T)] \log t]. \tag{3}$$

However, the dependence of $A(\epsilon, T) = F(\epsilon, T)$ upon $T$ was left open. The present paper fills in the gap and brings the fundamental stress-temperature relationship for cork.

### EXPERIMENTAL METHOD

The apparatus used was described in the previous paper. It consisted of a lever arm for compressing the sample which was automatically kept balanced through a chainomatic drive. Through this drive the force was recorded as a function of time on a clock-driven drum.

The apparatus was carefully rechecked for possible errors in the thermal expansion correction and was also corrected for rate of thermal expansion. This latter is very important in connection with the measurement of stress as a function of temperature at constant compression, since the compression must be maintained constant during the temperature change as well as during the temperature equilibrium. The procedure used was patterned after that of Anthony, Caston, and Guth.[2] That is, the sample was allowed to relax[1] until the rate of change of stress was small enough to make the stress-relaxation negligible during the subsequent stress-temperature determination. The preliminary relaxation time used was 1000 minutes. After this

---

[1] Dart and Guth, J. App. Phys. 17, 314–318 (1946).

[2] R. L. Anthony, R. H. Caston, and E. Guth, J. Phys. Chem. 46, 826 (1942). Reprinted in Rubber Chem. Tech. 16, 297 (1943).

preliminary relaxation period the temperature was lowered in steps and the corresponding stress changes were noted. Using this procedure one can find the temperature behavior of the stress at temperatures less than the relaxation temperature.

Figure 1 shows the results of such an investigation at the compressions indicated and with the common relaxation temperature of 80°C. It is seen that within the experimental error the stress depends linearly on the temperature. Thus, we may express this by the equation

$$Z(\epsilon, T) = M(\epsilon) + N(\epsilon)T, \tag{4}$$

where $M(\epsilon)$ and $N(\epsilon)$ are functions of compression alone. We will defer the discussion of the significance of this relation to a later section.

This then gives us the stress as a function of compression and temperature, exclusive of time effects. That is, this Eq. (4) gives us the temperature dependence of the factor $A(\epsilon, T)$ or $F(\epsilon, T)$ in Eq. (3). Before we write down an equation expressing all the experimental results presented, we must recognize that we have measured time effects and temperature effects separately, i.e., time effects were measured at constant temperature and *vice versa*. Let us call the constant relaxation temperature $T^*$ and the temperature which is changed, $T$. $T^*$ is really a variable separate from $T$ because in order to specify the state of the system one must specify not only the actual temperature $T$, but also the temperature $T^*$ at which the relaxation took place. This situation was discussed in more detail by Guth, Anthony, and Wack[3] for the case of rubber in extension. They showed, experimentally, that certain ratios (cf. Eqs. (13a, b, c, and d) of their paper) are constant. The constancy of these ratios implies the validity of "generalized similarity and separability" expressed by the equation

$$Z(\epsilon, T, T^*, t) = F(\epsilon, T) \cdot G(T^*, t). \tag{5}$$

It seems reasonable to assume that "generalized similarity" will also hold for cork, though this was not checked directly. Taking into account (3) we can then identify more explicitly the factors $F$ and $G$.

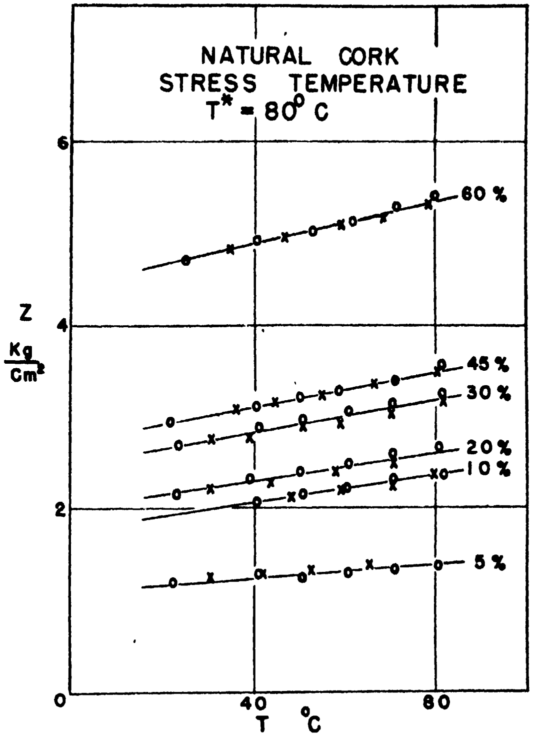


FIG. 1. Stress-temperature curves for cork previously relaxed at $T^* = 80°C$ for 1000 minutes. Circles indicate points obtained on lowering the temperature. Crosses are the points obtained on raising the temperature.

$$F(\epsilon, T) = M(\epsilon) + N(\epsilon) \cdot T;$$
$$G(T^*, t) = \{1 - [\beta/(\alpha - T^*)] \log t\}. \tag{6}$$

Then our generalized equation of state takes the form

$$\begin{aligned} Z(\epsilon, T, T^*, t) &= [M(\epsilon) + N(\epsilon)T] \\ &\quad \times \{1 - [\beta/(\alpha - T^*)] \log t\} \qquad (7) \\ &= F(\epsilon, T) \cdot G(T^*, t). \end{aligned}$$

## DISCUSSION OF EXPERIMENTAL RESULTS

### A. Stress Relaxation

The discussion so far has been concerned solely in presenting the experimental results of our investigation of the elastic properties of cork. We shall now try to give a more fundamental discussion of these results in the light of present day theories. Consider first, the stress relaxation. The general equation for the time dependency of stress as given by Boltzmann has been generalized for the non-Hookean case, by Guth, Wack,

---

[3] E. Guth, P. E. Wack, and R. L. Anthony, J. App. Phys. **17**, 347 (1946).

and Anthony.[3] This may be written as follows:

$$Z(t)=F[\epsilon(t)]-\beta'\int_{-\infty}^{t}\phi(t-t')F[\epsilon(t')]dt',\qquad(8)$$

where $\beta'$ is a small constant, $F=F(\epsilon)$ gives the stress-strain relationship, and $\beta'\phi(t-t')$ is the memory function of Boltzmann. For relaxation we have $\epsilon(t)=\epsilon(t')=0$ for $t<0$ and $\epsilon(t)=\text{constant}$ for $t>0$. Thus, Eq. (8) becomes

$$Z(t)=F(\epsilon)[1-\beta'\psi(t)],\qquad(9)$$

where

$$\psi(t)=\int_{0}^{t}\phi(t-t')dt'.\qquad(9a)$$

It is to be noticed that Eq. (9) is identical in form with the experimental Eq. (3). Thus, we can describe the time dependency by means of the generalized Boltzmann theory. Also, for the case of cork, comparison of (3) and (9) shows that

$$\beta'\psi(t)=[\beta/(\alpha-T)]\log t,\qquad(10)$$

while (9a) shows that

$$\phi(t-t')=1/(t-t').\qquad(10a)$$

A relation of the type (10) holds for many materials in a limited range of $t$ values. The remarkable result for cork is the validity of the relation (10) over a very wide range of $t$ values. As a matter of fact, (10) is valid for cork in the range $t=1$ min. We plan to discuss the theoretical significance of this fact at a later time.

If the stress-time curve is an exponential in $t$, we have

$$Z=Z_0\exp(-K't),\qquad(11)$$

where $K'$ is given by

$$K'=(kT/h)e^{-\Delta H/RT}.$$

Here $\Delta H$ is the activation energy and $kT/h$ is the rate factor as given by Eyring.[4] This equation is a straight line on a $\log Z$ *vs.* $t$ plot. Such a plot was tried on a relaxation run at 200°C and 4.86 percent compression and the result was far from linear. However, activation energies were calculated using a straight line portion of the curve and also using the time for the stress to drop to a value $1/e$ times the first value. These activation energies were 38.0 and 36.8 kilocalories per mole, respectively. These results are slightly larger than the corresponding values for rubber-like materials, which have activation energies of about 32 kilocalories per mole.

In Eq. (3), $F(\epsilon,T)$ designates the value of the stress at $t=1$ min. Expressing this fact more explicitly, Eq. (7) may be written

$$Z(\epsilon,T,T^*,t)=F(\epsilon,T)\times\{1-[\beta/(\alpha-T^*)]\log t/t_0\},\qquad(12)$$

where $t_0=1$ min. For $\alpha\leq T^*$ Eq. (12) was not established experimentally and it cannot be extrapolated to this range without further considerations. At any rate theoretical arguments show that a logarithmical relationship of the type of Eq. (12) will not hold for very small times, i.e., for times much smaller than 1 min.

In the first installment of this series[1] the time $\tau$ needed for the stress to decay to zero, for a sample held at constant compression, was shown to be independent of the compression strain. To show the dependence of $\tau$ upon the relaxation temperature $T^*$, a plot of $T^*$ *versus* $\log\tau$ was used. The curve so obtained was a straight line.

Instead of a $T^*$ *vs.* $\log\tau$ plot, a $\log\tau$ *vs.* $1/T^*$ plot may also be used. A straight line does not fit in the latter representation as well as the former one, but still gives a reasonable fit, if one considers the unavoidable experimental errors. We then have the relation

$$1/T^*=a+b\log\tau.$$

Solving for $\tau$ we obtain

$$\tau=\tau_0\exp(V/RT^*),\qquad(13)$$

i.e., an infinite decay time at absolute zero, which is reasonable. The constants in Eq. (13) have the values

$\tau_0=1.6\cdot10^{-3}$ min.,
$V=9$ kilocalories per mole.

### B. Stress-Temperature

Concerning the stress-temperature relationship, we can proceed as follows:

We start from the well-known energy equation of thermodynamics

$$\left(\frac{\partial U}{\partial V}\right)_T=T\left(\frac{\partial P}{\partial T}\right)_V-P.\qquad(14)$$

---

[4] Cf. for instance, S. Glasstone, K. Laidler, and H. Eyring, *The Theory of Rate Processes* (McGraw-Hill, New York, 1941).

Using the Maxwell relation

$$\left(\frac{\partial P}{\partial T}\right)_V=\left(\frac{\partial S}{\partial V}\right)_T, \tag{15}$$

we can write

$$\left(\frac{\partial U}{\partial V}\right)_T=T\left(\frac{\partial S}{\partial V}\right)_T-P. \tag{16}$$

Applying this to our case of unilateral compression we must substitute $1-\epsilon$ for $V$, and $Z$ for $P$, since we plot actually the pressure or the negative of the tension stress. This gives

$$-\left(\frac{\partial U}{\partial \epsilon}\right)_T=-T\left(\frac{\partial S}{\partial \epsilon}\right)_T-Z$$

or

$$Z=\left(\frac{\partial U}{\partial \epsilon}\right)_T-T\left(\frac{\partial S}{\partial \epsilon}\right)_T=Z_U+Z_S, \tag{17}$$

where $Z_U$ is the contribution of the internal energy to the stress, $Z_S$ is the entropy contribution, and $Z$ is, of course, the total stress. (Actually the experiments are carried out at constant pressure and constant compression. Nevertheless, Eq. (17) is valid in sufficient approximation, when derivatives are taken at constant temperature and constant pressure.) Equation (17) has the same form as our experimental Eq. (4). Thus, using the curves shown in Fig. 1 we can calculate $Z_U$ and $Z_S$ as a function of $\epsilon$. In doing this one must remember that the $T$ in Eq. (17) is absolute temperature. The results of such a calculation are shown in Fig. 2. Here one can see that the entropy is responsible for most of the stress in the first two regions, i.e., for the stiffness of the sample at small compressions and during the breakdown of the structure. However, when the actual material of the cell walls begins to be compressed, at 60 percent or so, the internal energy contribution becomes appreciable.

The negative internal energy contribution in the first range of (small) compressions is at least partly caused by the thermal expansion of the uncompressed cork. This is similar to the behavior of rubber at small extensions. The internal energy contribution is partly caused by the thermal expansion of the unstretched rubber.

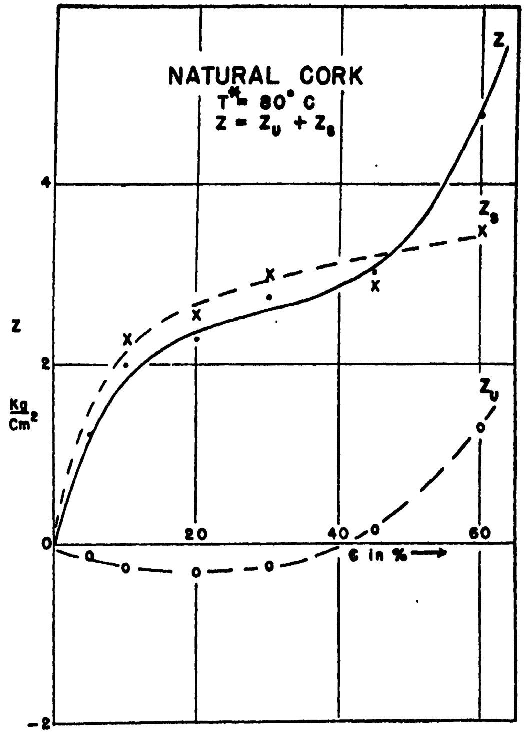


FIG. 2. Resolution of the stress of Fig. 1 at $T=25°C$ into its components $Z_U$ and $Z_S$ due to internal energy and entropy, respectively.

The large contribution to the stress due to entropy is presumably connected with the bending of the cell walls. The material of the cell walls is likely to consist of long chain molecules with at least some degree of free rotation. On this basis a large entropy contribution can be understood. The increase of internal energy, on further compression, is presumably caused by intermolecular forces coming into prominence as compression increases.

## ACKNOWLEDGMENT

The authors wish to express their appreciation to Dr. H. A. Robinson, Chief Physicist, Armstrong Cork Company, for valuable discussions and for furnishing the samples used in this investigation. They also wish to express their appreciation to Mr. E. Claxton, Director of Research, and to Dr. R. B. Rohrer, Assistant Director of Research, Armstrong Cork Company, for the benefit of valuable discussions and for the support of this work.

# Elastic Properties of Cork

## III. Hydrostatic and Ordinary Load-Compression Curves for Cork

S. L. DART,* HOWARD A. ROBINSON,** AND EUGENE GUTH*

(Received October 14, 1946)

Load-compression curves were taken for natural cork. The S-shape of the curves is discussed and a theoretical fit is made for the high end of the curve, assuming an initial stiffness and an incompressible part of the cork material. A comparison is made between repeated compressions, with and without a recovery period of time.

Pressure-volume curves were also taken and their significance discussed in connection with the load-compression curves leading to an evaluation of Poisson's ratio.

DETERMINATION of the load-compression curves of cork under various experimental conditions is of great theoretical and practical interest. The load-compression curves for cork show a peculiar S-shape. The proper interpretation of this S-shape should throw considerable light on the structure of cork. A knowledge of the structure of cork, in turn, will facilitate the design of corky materials for practical applications, and enable one to predict their behavior under a great variety of conditions.

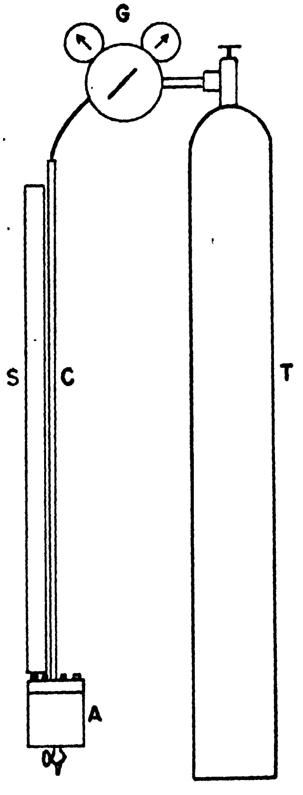


FIG. 1. Apparatus for hydrostatic compression of cork. *T*: high pressure tank; *G*: gauge and reducing valve; *S*: scale; *C*: glass capillary tube; *A* : compression chamber.

In this paper an attempt is made to interpret the S-shaped load-compression curve of cork as the result of the cellular structure and of the two-phase structure of cork, namely, the corky material *and* the air trapped in the cork cells. Load-compression curves were obtained in two ways: in the ordinary way, and second, by measuring the change in volume caused by hydrostatic pressure.

These two methods lead actually to somewhat different load-compression curves; their mutual relation will be explained in this paper in the section on hydrostatic curves.

### DESCRIPTION OF THE EXPERIMENTAL METHOD

#### A. Ordinary Load-Compression Curves

These curves were obtained by means of the apparatus described in a previous paper[1] (cf. Fig. 1 there). The same sample was used to obtain the complete load-compression curves up to 70 percent compression by gradually increasing the load, and reading the corresponding compressions on the dial gauge.

#### B. Hydrostatic Volume-Pressure Curves

These curves were obtained on an experimental arrangement shown in Fig. 1. A cork

---

* Polymer Physics Laboratory, University of Notre Dame, Notre Dame, Indiana.

** Research Laboratories, Armstrong Cork Company, Lancaster, Pennsylvania.

[1] Dart and Guth, J. App. Phys. 17, 314 (1946).

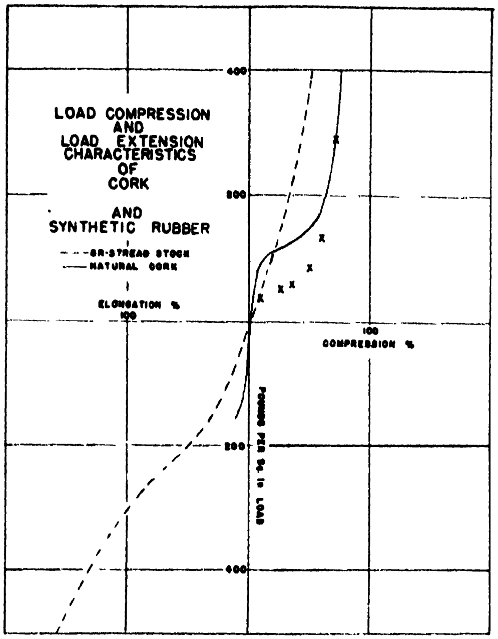


FIG. 2. Load compression and extension characteristics of cork and of GR-S tread. The crosses indicate the curve obtained by repeating the loading a second and a third time on the same sample.

sample of about six cubic centimeters volume is placed in the compression chamber which is then bolted shut. Water is then forced into the stopcock at the bottom until it rises near the top of the glass tube and scale. The stopcock is then closed and any air bubbles are made to rise to the top of the tube by tapping. Care must be taken to get all air bubbles out as they would cause an error in the volume-pressure curve. A zero reading on the scale is taken of the water level, and the whole unit is connected through a reducing valve to a high pressure tank (of nitrogen in our case). The pressure is then increased in units of 5 p.s.i. every 30 seconds, and the corresponding water-level readings on the scale are taken at the end of each 30-second period. This procedure was used in order to simulate as closely as possible the procedure used in taking an ordinary load-compression curve. The pressure was read from the reducing gauge and the volume change was calculated from the scale readings (knowing the diameter of the glass tube). It was interesting to note that there was a great deal of flow of the cork during a run, as is the case during an ordinary load-compression run. This implies that the flow is caused by the cork structure itself and not by the escape of air from the cells, which would be impossible in the hydrostatic case.

## ORDINARY LOAD-COMPRESSION CURVES FOR CORK

Figure 2 represents load-compression curves obtained at the Armstrong Research Laboratories. Up to the breaking point the extension branch of this curve is also given for cork. The crosses indicate the load-compression curve for cork obtained by repeating the loading a second and a third time on the same sample. Twenty-four hours elapsed between consecutive loadings.

Figure 3 shows a somewhat more extensive study of the repeated load-compression curves carried out at Notre Dame. The consecutive curves were taken one after the other without giving time for the sample to recover. Complete hysteresis loops were obtained.

Figures 2 and 3 reveal a series of interesting features important for the interpretation of the load-compression curves.

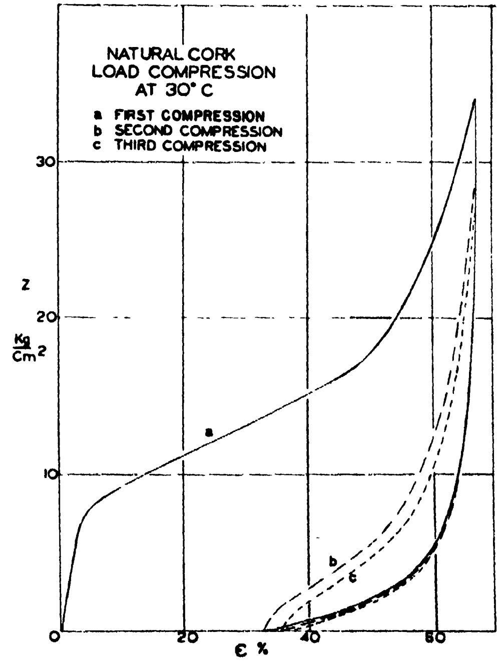


FIG. 3. Load-compression curves for repeated compressions including hysteresis loops.

The extension branch is monotonic. The tensile strength and breaking elongation are low. They

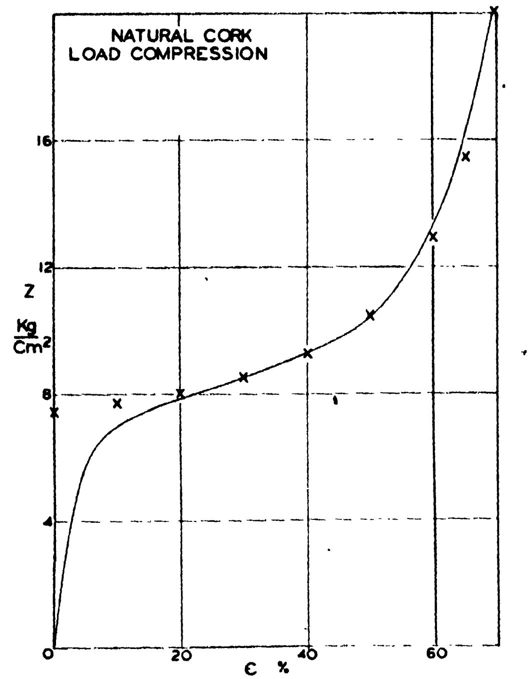


FIG. 4. Load-compression curve of cork. Crosses indicate theoretical Eq. (1) of text.

are chiefly determined by the cuts or cracks invariably present in cork, and the propagation or growth of these cracks. The tensile strength may be compared with the tear resistance of a compact material like rubber, the tear being natural and unavoidable for cork.

The compression branch shows a characteristic S-shape. On repeated compression the S-shape tends to be washed out and the curve approaches somewhat the monotonic behavior shown by GR-S tread.[3] This is shown with particular clarity in Fig. 3, but to a lesser degree in Fig. 2 because of the recovery of the sample between consecutive loadings.

A simple explanation of the S-shape may be based on the known cellular and two-phase structure of cork: the corky material (including various organic substances) constituting the cell walls and the air trapped in the cells. We notice that the knee of the compression curve occurs roughly at a compression corresponding (on inversion at the origin) to the breaking elongation, and for a load somewhat smaller than the negative of the tensile strength. On repeated stretching the knee occurs for a lower critical load. The equal order of absolute magnitude of the critical load and compression of the knee and of the tensile strength suggest that a certain breakdown of the cells may have taken place at the knee. The difference in the lowering of the knee as shown in Fig. 2 with recovery period between repeated loadings, and in Fig. 3 without a recovery period, indicates that the breakdown of the cells is partially reversible.

For the part above the knee of the compression curve we have tried a preliminary quantitative representation. That part of the compressive stress $Z$ may be represented fairly accurately by the equation

$$(Z-a)(L-l_0)=\text{const}; \qquad (1)$$

$L$ is the relative length, $a$ and $l_0$ are constants. The quantity $a$ should be independent of the compression, and represents the critical stress necessary for the partial destruction of the cork structure (*not* the *cells*).

The plateau after the knee is interpreted as largely caused by the failure of the resistance of the cork (through buckling of the cell walls for instance). The upward turn is interpreted as a compression of the almost incompressible part of

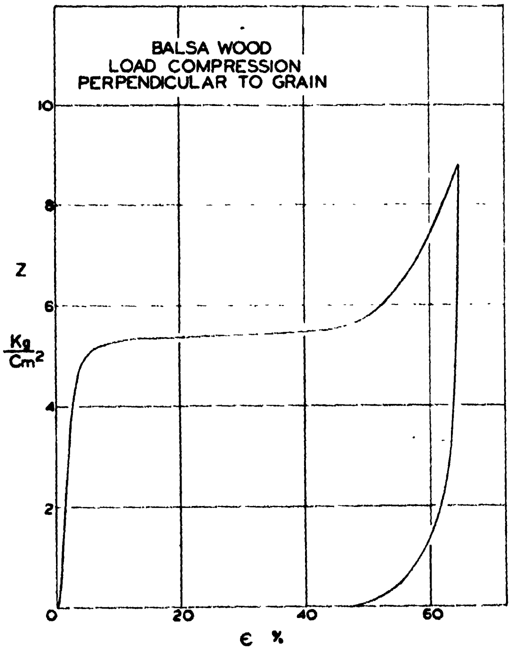


FIG. 5. Load-compression loop for balsa wood.

---

[3] Synthetic rubber compound containing carbon black as used for the tread surface of tires.

cork. Of course, the van der Waals $b$ of the air comes in also. The constant on the right side of Eq. (1) should be of the order $RT$. According to this equation, $Z$ should be a linear function of the temperature at constant compression (since the constant on the right side of Eq. (1) should be proportional to $T$)

$$Z = M(L) + N(L) \cdot T. \qquad (2)$$

This conclusion was checked in the second paper of this series by our relaxation method. Equation (1) checks well with a load-compression curve, as shown in Fig. 4. However, the true significance of this apparent agreement has not been established. For some samples the plateau above the knee was practically straight. The possible significance of this finding, at some variance with Eq. (1), needs further study.

It is interesting to note that balsa wood, when compressed perpendicular to the grain, also shows a pronounced S-shaped load-compression curve, with an almost horizontal plateau, as shown in Fig. 5. The large set (46 percent) after compression disappears almost completely when the sample is soaked in water.

### LOAD COMPRESSION CURVE OF HIGHLY COMPRESSED CORK

In connection with the interpretation of the S-shaped load-compression curve it was interesting to see how a high hydrostatic pressure would change the shape of the load-compression curve. At our request Dr. P. W. Bridgman[3] was kind enough to apply to cork a pressure of 29,000 kg/cm² in isopentane. The initial dimensions of the sample on April 11, 1945, were 0.324″×0.319″ ×0.312″. After exposure to high pressure the dimensions were 0.297″×0.301″×0.283″. On February 6, 1946, the dimensions were 0.302″ ×0.306″×0.287″. In view of the magnitude of the hydrostatic pressure employed in the test, it is remarkable that the permanent set of the linear dimensions was only about 8 percent on the average immediately after the exposure. After ten months' rest the permanent set was somewhat less.

In Fig. 6 the dotted line shows the load-compression curve taken in June, 1946, on the same piece of cork which had been subjected to the high pressure by Dr. Bridgman on April 11, 1945. The full line shows the load-compression curve of Fig. 2, which was obtained on a piece of cork not previously compressed. It is seen that the initial part of the two curves is almost the same. We interpret this initial part as resulting from the initial bending of the cell walls. The comparison of the two curves exhibits nicely the fact that the cells were not crushed by the high hydrostatic pressure. The plateau of the curve for the pre-compressed sample is, however, much less developed than in the case of the previously uncompressed sample. We attribute this plateau to the failure of the resistance of the cork cells. The comparison of the two curves shows that the hydrostatic compression must have driven a large amount of the air out of the cells. The steep part of the two curves has about the same slope, as one would expect.

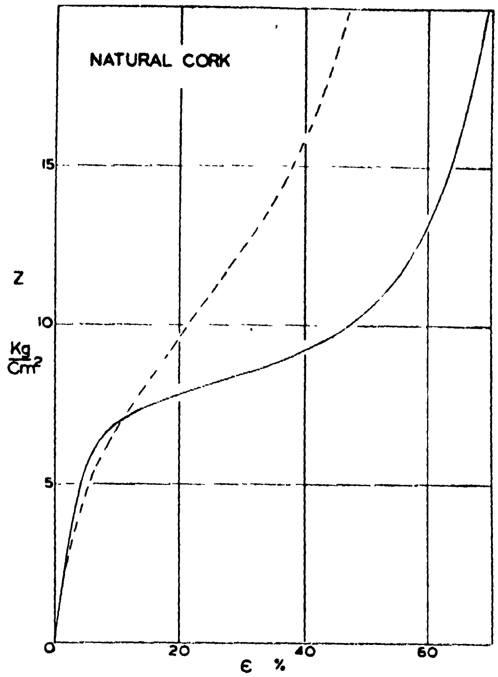


FIG. 6. Comparison of load-compression curves for cork; dashed line shows curve for sample hydrostatically compressed by P. W. Bridgman, applying a pressure of 29,000 kg/cm². Full line shows behavior of a sample which was not previously compressed.

### HYDROSTATIC CURVES

It is clear from the description of this method in the first section that hydrostatic compression

---

[3] Jefferson Physical Laboratory, Harvard University, Cambridge, Massachusetts.

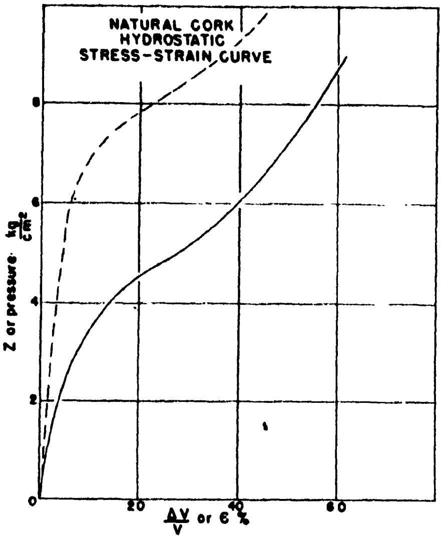


FIG. 7. Comparison of hydrostatic (full line) and ordinary load-compression (dashed line) curves for cork.

leads to a pressure-volume relationship, different from the ordinary load-compression relation. The connection between these two types of relations may be exhibited simply and quantitatively, referring to the elements of the theory of elasticity. The slope at the origin of a load-compression curve $Z=Z(\epsilon)$ is Young's modulus

$$dZ/d\epsilon=E. \tag{3}$$

Similarly, the slope at the origin of the pressure-volume relation, $P=P(V)$, is the bulk modulus $K$

$$VdP/dV=K. \tag{4}$$

Now according to the theory of elasticity, Young's modulus, $E$, and the bulk modulus $K$ are connected as follows:

$$E=3K(1-2\mu), \tag{5}$$

where $\mu$ is Poisson's ratio (ratio between lateral contraction and longitudinal extension). Knowing $E$ and $K$ Eq. (5) enables us to obtain $\mu$.

$$\mu=\tfrac{1}{2}(1-E/3K). \tag{6}$$

Figure 7 shows a hydrostatic pressure-volume change curve with the corresponding load-compression curve for cork. When inserted into Eq. (6), comparison of the slopes of the two curves yields

$$\mu=0.04. \tag{7}$$

Another method of obtaining $\mu$ is to measure the change of volume on ordinary compression. According to the theory of elasticity the following relation holds:

$$\Delta V/V=\epsilon(1-2\mu). \tag{8}$$

This relation permits one to compute $\mu$ if $\Delta V/V$ is measured for a given compression $\epsilon$. This method, however, was not used in the present work.

## ACKNOWLEDGMENT

The authors wish to express their appreciation to Mr. E. Claxton, Director of Research, and to Dr. R. B. Rohrer, Assistant Director of Research, Armstrong Cork Company, for the benefit of valuable discussions and for the support of this work.

# Contributed Original Research

## A Cold Cathode Rectifier

WILLARD H. BENNETT*

(Received October 21, 1946)

A cold cathode rectifier is described in which corona discharge is used at atmospheric and higher pressures in hydrogen and nitrogen free of electron-attaching impurities. This kind of tube has advantages where the current demands upon a rectifier are small.

IN the course of some experimental studies of corona discharge, methods were found for producing and maintaining conditions for free electron discharge in gases at pressures of the order of magnitude of atmospheric and higher. The very large difference in mobility between free electrons and positive ions is illustrated in Fig. 1, in which are shown the negative and positive corona current-voltage characteristics for a 1-mil tungsten wire-end discharging in hydrogen at atmospheric pressure towards a plane electrode at a distance of 3 cm. Characteristics for wire sizes between 0.4 and 3 mil, and for distances to the plane electrode between 0.3 cm and 10 cm were measured and are summarized in an earlier report.[1]

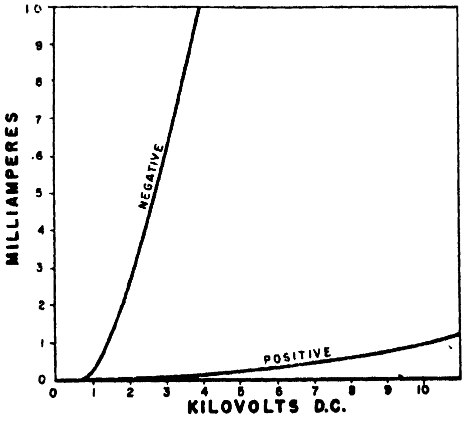


FIG. 1. Current-voltage characteristics for negative and positive corona.

As will be seen from the figure, the very large difference between the mobilities of electrons and positive ions is quite unusual for corona discharges. These can only be attained by using means for eliminating electron-attaching components from the gas in the tube to a degree estimated as less than one part per hundred million and keeping the gas free of these components to this degree.

Cold cathode rectifiers were developed taking advantage of this large difference in mobilities. It is believed that these rectifiers have peculiar advantages over conventional types of rectifiers in many applications, particularly where the

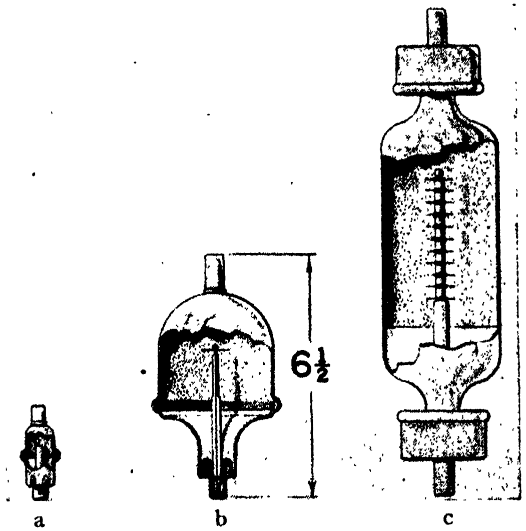


FIG. 2. Representative types of tube. (The dimension 6½ is in inches.)

---

[*] Now at the National Bureau of Standards. This work was performed in 1939–1941 at the Electronic Research Corporation, Newark, Ohio.

[1] Willard H. Bennett, Phys. Rev. 58, 992–997 (1940), see Figs. 2 and 3, p. 994. The abscissae in those figures were incorrectly labelled and should have been "kilovolts difference" instead of "log voltage difference."

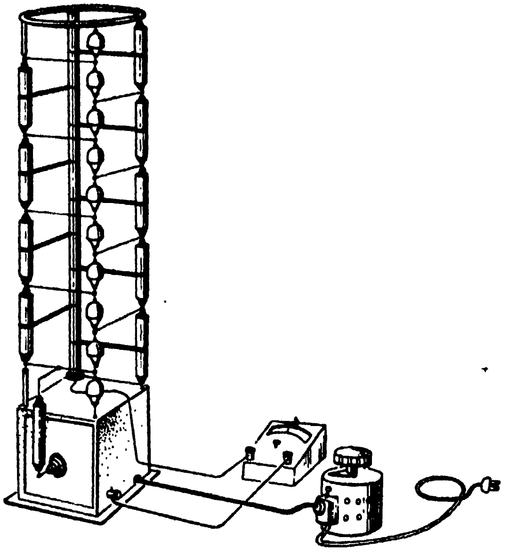


FIG. 3. A representative type of rectifier using ten tubes of the type shown in Fig. 2b. The condensers are 0.5 microfarad each and rated at 20 kilovolts d.c. each. The multiplier can be a spiral ink line resistor on a Bakelite or formica cylinder with 100 megohms per stage.

current requirements are small, and that a description of this kind of rectifiers would be of interest.

## GENERAL DESCRIPTION

Some representative types of tube are illustrated in Fig. 2. In these tubes, the discharge electrodes are one- or two-mil tungsten wire-ends. The collecting electrode in each is an Aquadag coating on the glass wall shown as the shaded area in the broken sections.

The tubes in Figs. 2a and 2b are filled with hydrogen at atmospheric pressure. The tube in Fig. 2c is filled at eight atmospheres with 30 percent hydrogen and 70 percent nitrogen.

Because there is no filament heating requirement with a tube of this kind, rectifiers built using these tubes are simpler, lighter, and more compact, than rectifiers using conventional types of tubes. As an illustration, Fig. 3 shows a rectifier supplying 100 kilovolts using a 10 kilovolt transformer in a Cockcroft-Walton circuit. In this kind of circuit, if conventional types of tubes were used, the need for filament heating current at each stage would make necessary a filament transformer with high voltage insulation between primary and secondary for each stage. Such a requirement would make necessary a bulkier and more complicated equipment than that shown in Fig. 3.

There is no requirement for filament heat-up time and no devices are needed for tube protection on this account.

Rectifiers can readily be designed with clips for holding the tubes so that simply reversing the tubes in the clips reverses the output polarity of the rectifier.[2]

Although these tubes do not waste electric power in heating a filament, they suffer from loss of efficiency from another cause. Under normal conditions of operation, there is a small positive-ion current on the reverse half-cycle. This effect together with the fact that even the high free electron mobilities are not sufficient to give as high current densities as are obtainable with electrons in vacuum, has served to hold these tubes to an efficiency of about 20 percent to 50 percent which is lower than most conventional types of rectifier unless no greater output currents are required than can be adequately supplied by these cold cathode tubes.

In Figs. 4 and 5 are shown some efficiency data taken on a rectifier using the type of tube shown in Fig. 2c. In Fig. 4, the transformer supplies 10 kilovolts a.c., and in Fig. 5, the supply is 20 kilovolts a.c. Observations were obtained for hydrogen (shown with full lines) and a mixture of 30 percent hydrogen and 70 percent nitrogen (shown with dashed lines). War work prevented completing these data, but from some observations in a single wire experimental tube, a good estimate was obtained for 20 percent deuterium and 80 percent nitrogen (shown with the dotted lines). From mobility data[3] it is to be expected that the efficiency would be still higher for a mixture of 5 percent deuterium and 95 percent nitrogen. For comparison purposes, representative curves for a vacuum-tube rectifier are included.

## PREPARATION OF A TUBE

As discharge electrodes, wire-ends with any diameter between about 0.4 mil and 3 mils seem

---

[2] Additional engineering data is given in Trans. A.I.E.E. 64, 164–7 (1945); and Electronics Industries, p. 106 (November, 1945).

[3] W. H. Bennett and L. H. Thomas, Phys. Rev. 62, 41–47 (1942).

to be about equally effective. The reason for this is that in the initial conditioning of the tube, there is an etching of the wire-ends to a sharp conical tip whose radius of curvature is of the order of magnitude of 0.2 mil.[4] In tubes containing six or more wire-ends, the maximum number that can usefully be grouped with bases adjoining is six, and even with six wires in a group, the maximum current per wire-end that can be drawn without arc-over is less than six times the current that could be drawn from a single wire, the other electrode spacings being the same but the other five wires removed. The discharge spreads laterally from a wire-tip and adjacent discharges interfere with each other to some extent.[5]

It is important that the electric field diverge all the way to the collecting electrode. Aquadag or metal collecting electrodes can be used.

After filling the tube with gas and sealing the tube, a.c. voltage is applied directly. The resulting discharge eliminates the electron-attaching impurities and sharpens the wire-ends, slowly at first. The current rises quite rapidly during the last stages of the conditioning, however, and unless a circuit breaker is used in the a.c. supply, the tube may be greatly over-heated.

This does not complete the conditioning. If the tube is allowed to stand, it will probably decondition to some extent after a day, that is, the electron mobilities will have decreased by a large amount. This is probably caused by the liberation into the gas of some electron-attaching impurities from the walls of the tube and from the electrodes. For this reason, immediately following the first run-up, the tube is put on a reduced a.c. voltage sufficient to maintain a current of about 100 microamperes per wire. After about three hours, the tube is permanently conditioned.

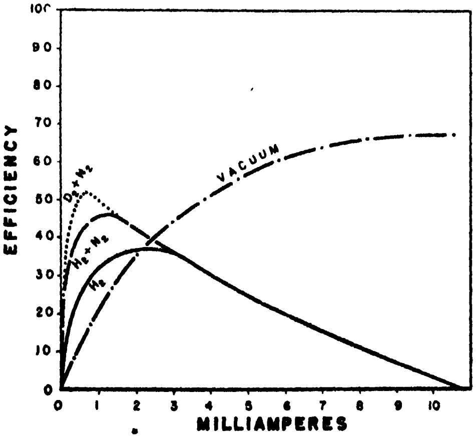


FIG. 4. Efficiency with 10 kv a.c., applied.

## OPERATION

One of the convenient features of this kind of tube is that the full rated voltage can be thrown on without any preliminaries.

There is a definite output voltage at which a tube arcs over and the maximum d.c. output rating of the tube can be given as about 10 percent under that value. The tube never "arcs-back" at unexpectedly low voltages as some of the conventional mercury vapor tubes do occasionally.

There is no observable consumption of discharge electrodes or other materials in these tubes and there is no known reason for these tubes to have other than an indefinitely long life when operated within the ranges of rated current and voltage.

Surges in the high voltage a.c. supply, caused by opening or closing the primary circuit, cannot puncture the tube with cold emission because the tube is filled with gas.

The small positive-ion current on the inverse half-cycle serves a useful purpose. When the output of the rectifier is being smoothed with a condenser, resetting the input a.c. voltage to a

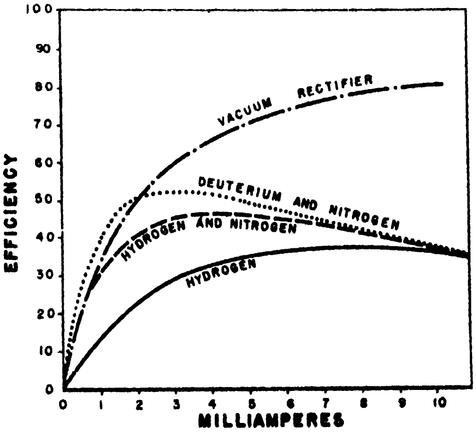


FIG. 5, Efficiency with 20 kv a.c., applied.

---

[4] Willard H. Bennett, Phys. Rev. 61, 55 (1942). See also reference 1.

[5] Willard H. Bennett, Phys. Rev. 61, 53–56 (1942).

lower value results in the condenser voltage dropping correspondingly because of the positive-ion current in the tube.

As long as the rated inverse voltage is not exceeded with these tubes, no damage results from drawing more than the rated current through the tubes momentarily. The only damage that eventually occurs is caused by over-heating the tube as a whole and this takes at least several minutes for all types of tubes. The largest tubes, containing part nitrogen, decondition at currents of about 0.2 milliampere per wire because of a slow formation of ammonia. Operation at 0.06 milliampere per wire slowly disassociates the ammonia and reconditions the tube.

At the higher values of current, ionization by collision occurs between the sheath and the collector in a tube and a high pressure glow discharge makes its appearance. This glow discharge has a positive resistance characteristic for currents up to a higher order of magnitude than the limiting current for the corona discharge, and is not to be confused with normal high pressure arcs. This is not a disruptive kind of arc (that is, it is not abruptly initiated with a streamer and a spark) because of the absence of electron-attaching gaseous components.[6]

In tubes containing part nitrogen, at currents of the order of ½ milliampere per wire-end, the wire-ends become heated, and at currents of between 1 and 5 milliamperes per wire-end, the wire-ends attain temperatures as high as 2000°C. The energy necessary to produce these temperatures cannot be accounted for by the impact energies of positive ions striking the wire-ends following their last free paths. Such heating does not occur in discharges in pure hydrogen at the same currents.[7] It appears that either an active state of nitrogen, or else an excessive concentration of atomic hydrogen caused by the presence of nitrogen, is produced in the ionization sheath and is responsible for this excessive heating which occurs only in part nitrogen.

When the wire-ends heat to more than about 1000°C, a reduction in sheath drop in potential[1] occurs both for positive and negative discharge because of the thermal reduction in gas density immediately adjacent to the wire-end. There is a corresponding reduction in arc-over voltage for the tube on the inverse half-cycle. This effect is reduced by using the finest wires (0.4 mil) but is not entirely overcome by this means.

## CONCLUSION

It is not believed that the full potentialities of this kind of discharge in a rectifier have been realized. Further studies of the diffuse kinds of arcs which occur in these gases should prove profitable. It is thought that a wider appreciation of the advantages of a cold cathode gas-filled rectifier may bring an increased interest in studies of the properties and methods for control of these arcs.

---

[6] Willard H. Bennett, J. App. Phys. 13, 199–200 (1942). William H. Otto and Willard H. Bennett, J. Chem. Phys. 8, 899–903 (1940).

[7] Willard H. Bennett, Phys. Rev. 62, 369–371 (1942).

---

# The Hydrodynamic Lubrication of Finite Sliders

Charles P. Boegli
*Springfield, Ohio*

(Received December 17, 1946)

Two approximations are made in the solution of Reynolds' lubrication equation for the case of a finite slider. These approximations lead to a series of equations that are easy and rapid to use for flat sliders, and which are also applicable to curved sliders. Examples of their use are presented, and a number of calculations are made to determine the range of slider proportions to which they may be applied with sufficient accuracy. The method is of special value for wide sliders which are difficult or impossible to analyze by other methods.

## INTRODUCTION

THE solution of the basic Reynolds differential lubrication equation for the case of a slider bearing of finite width has always been beset with great difficulties. Reynolds[1] did not succeed in finding a general solution to the

---

[1] O. Reynolds, "On the theory of lubrication and its application to Mr. Beauchamp Tower's experiments," Phil. Trans. Roy. Soc. 177, part 1, 157–234 (1886).

equation, and it was not until nineteen years after his presentation that Michell[2] provided a satisfactory attack to the problem.

Michell's method involves no assumptions that affect the accuracy of the solution for the specified conditions. It is, however, applicable only to sliders with flat surfaces, and even for such simple cases the solution is so tedious that only a very small amount of data has yet been computed. R. O. Boswall, working with an assumption credited to Stodola, succeeded in developing an approximate method[3] that gives results negligibly different from those of the exact method in cases where comparisons are possible. Here again the analytical solution can only be made for flat sliders; for other shapes resort is had to a graphical method that actually consists in estimating by eye the shape of one of the pressure distribution curves.

More recently, Muskat, Morgan, and Meres published another solution to the original differential equation.[4] They presented their solution as a series of graphs that serve better for the solution of problems of operation than problems of design. Again, only bearings with flat surfaces are considered.

There is great need for a set of equations that will provide, with a minimum of effort, a good approximation to the actual behavior of a wider variety of sliders than can now be handled. The need for a very high degree of accuracy is, for several reasons, not great. One reason is that it is never possible to predict closely the behavior of the viscosity of the lubricant as it passes through the bearing; the assumption of constant viscosity is generally made, and the practical value of the solution is limited by this approximation. Also, it would be very difficult actually to construct a bearing in exact conformity with a specified design; such a small difference, for example, as a slight rounding of the leading edge, would have considerable effect upon the performance. Therefore, a solution that predicts to a fair degree of accuracy the performance of a great many types of slider bearings would be at least as valuable as one that predicts very closely the performance of only a narrow range.

Such a solution is provided through two simple assumptions. The first is that used by Boswall in his derivation of an approximate method; namely, that the pressure functions along the length and width of the slider are independent. The second is that the pressure function along the length of the slider (that is, the dimension in the direction of motion) is the *same as that for an infinite slider.*

This paper presents analyses based upon these two assumptions; an analytical solution for the case of the flat slider and an arithmetical method for curved sliders. Since with any approximate method some estimate of the accuracy and range of applicability is desirable, the report also presents a comparison of the results obtainable with the new method to any available data found in other manners.

## ANALYSIS

The general differential equation of Reynolds serves as the starting point for this analysis of slider lubrication.

$$\frac{\partial}{\partial x}\left(\frac{1}{\mu}\frac{\partial P}{\partial x}h^3\right)+\frac{\partial}{\partial z}\left(\frac{1}{\mu}\frac{\partial P}{\partial z}h^3\right)=6\frac{d}{dx}Uh. \quad (1)$$

In this equation $\mu$, $h$, $U$, and $P$ are the oil viscosity, the oil-film thickness, the slider velocity, and the oil pressure at any point on the slider surface. If the substitutions are made:

$$x=x_0+Lx_1,$$
$$z=Bz_1,$$
$$h=h_0\beta,$$

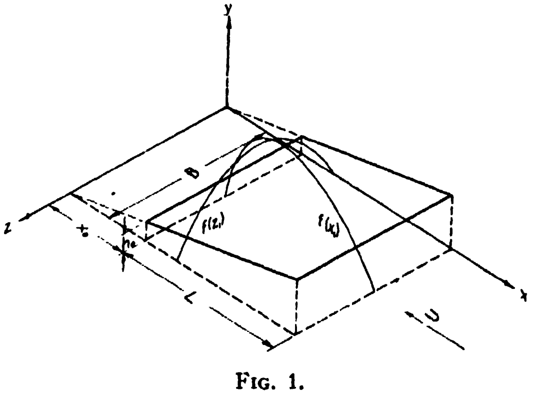


FIG. 1.

---

[2] A. G. M. Michell, "The lubrication of plane surfaces," Zeits. f. math. Physik **50**, 1073–1099 (1904).

[3] R. O. Boswall, *The Theory of Film Lubrication* (Longmans, Green, and Company, New York, 1928), pp. 141–158.

[4] Muskat, Morgan, and Meres, "Studies in lubrication. VII. The lubrication of plane sliders of finite width," J. App. Phys. **11**, 208–219 (March 1940).

where $L$ and $B$ are the length and width of the slider, $h_0$ represents the oil-film thickness at the outlet edge, and $\beta$ is a function of $x_1$ (see Fig. 1); and if viscosity is assumed constant, the equation becomes

$$\frac{\partial}{\partial x_1}\left[\beta^3\frac{\partial P}{\partial x_1}\right]+N^2\beta^3\frac{\partial^2 P}{\partial z_1^2}=A\frac{d\beta}{dx_1}. \tag{2}$$

In Eq. (2), $N$ is the ratio of slider length to slider width ($L/B$) and $A$ is the product $6\mu UL/h_0^2$ containing all the dimensional terms. The new variables $x_1$ and $z_1$ are the fractional distances along the length and width of the slider, and range in value from 0 to 1.0.

It will now be assumed, following the method in reference 3, page 141, that the pressure at any point can be expressed as the product of two functions

$$P=A\cdot f(x_1)\cdot f(z_1). \tag{3}$$

This product is substituted for $P$ in Eq. (2), yielding

$$\frac{\partial}{\partial x_1}\left[\beta^3 f(z_1)\frac{df(x_1)}{dx_1}\right]+N^2\beta^3 f(x_1)\frac{d^2f(z_1)}{dz_1^2}=\frac{d\beta}{dx_1}. \tag{4}$$

To find the form of $f(z_1)$ Eq. (4) is solved along the line where $f(x_1)$ is a maximum; that is, where $df(x_1)/dx_1$ is zero. This yields

$$\frac{d^2f(z_1)}{dz_1^2}+\frac{\left[\frac{1}{f(x_1)}\frac{d^2f(x_1)}{dx_1^2}\right]_{x_m}}{N^2}f(z_1)=\frac{\left[\frac{1}{\beta^3 f(x_1)}\frac{d\beta}{dx_1}\right]_{x_m}}{N^2}, \tag{5}$$

in which $x_m$ indicates the value of $x_1$ for which $f(x_1)$ is a maximum. This equation can be simplified by considering that since $f(x_1)$ is the same as that for an infinite slider, this being the second basic assumption, this function was derived from the differential equation

$$\frac{d}{dx_1}\left(\beta^3\frac{dP}{dx_1}\right)=A\frac{d\beta}{dx_1},$$

which, upon substitution of $P=A\cdot f(x_1)$ and performance of the indicated differentiation, gives

$$3\beta^2\frac{df(x_1)}{dx_1}+\beta^3\frac{d^2f(x_1)}{dx_1^2}=\frac{d\beta}{dx_1}.$$

Where $f(x_1)$ is a maximum, this equation shows that

$$\left[\frac{d^2f(x_1)}{dx_1}\right]_{x_m}=\left[\frac{1}{\beta^3}\frac{d\beta}{dx_1}\right]_{x_m} \tag{6}$$

and consequently Eq. (5) is reduced to

$$\frac{d^2f(z_1)}{dz_1^2}+\frac{M}{N^2}f(z_1)=\frac{M}{N^2}$$

where $M$ is equal to the expression on either side of Eq. (6) divided by $f(x_m)$. The solution is

$$f(z_1)=1-\left(\frac{e^n-1}{e^n-e^{-n}}\right)(e^{-nz_1}+e^{-n(1-z_1)}), \tag{7}$$

where

$$n^2=-\frac{M}{N^2}.$$

Up to this point the treatment has been general, and applicable to both flat and curved sliders. In the case of flat sliders,

$$\beta=1+bx_1$$

and

$$M=-\tfrac{1}{2}\frac{(b+2)^4}{(b+1)^2}$$

while $f(x_1)$ is simply[5]

$$f(x_1)=\frac{bx_1(x_1-1)}{(b+2)(bx_1+1)^2}. \tag{8}$$

For curved sliders, an arithmetic or graphical evaluation of $f(x_1)$ is usually preferred to an analytical solution. Such methods result in a series of tabulated values of the function for various values of $x_1$, and it is necessary to find $x_m$ and $f(x_m)$ from this table. An excellent method of finding the required values is presented in *Empirical Formulas.*[6] To evaluate $n$, the value of $M$ must be known; it can be calculated from the expression on the right side of Eq. (6), which is to be preferred to the quantity on the left, because the former does not require the finding of second derivatives from graphical data.

The important equations and constants are

---

[5] Reference 3, page 44.

[6] T. R. Running, *Empirical Formulas* (John Wiley and Sons, Inc., New York, 1917), pp. 100–113.

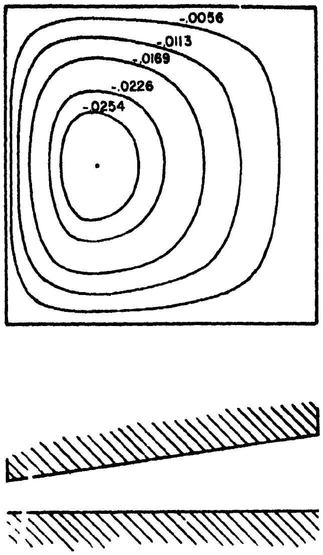


FIG. 2.

summarized below:

$$P = A \cdot f(x_1) \cdot f(z_1) \tag{3}$$

$$f(z_1) = 1 - \left( \frac{e^n - 1}{e^n - e^{-n}} \right) (e^{-nz_1} + e^{-n(1-z_1)}) \tag{7}$$

$$f(x_1) = \frac{bx_1(x_1 - 1)}{(b+2)(bx_1+1)^2} \quad \text{for flat sliders only} \tag{8}$$

in which

$$A = \frac{6\mu UL}{h_0^2}$$

$$n^2 = -\frac{M}{N^2}$$

$$M = \left[ \frac{1}{f(x_1)} \frac{d^2 f(x_1)}{dx_1} \right]_{x_m} = \left[ \frac{1}{\beta^3 f(x_1)} \frac{d\beta}{dx_1} \right]_{x_m}$$

$$= -\tfrac{1}{2} \frac{(b+2)^4}{(b+1)^2} \quad \text{for flat sliders only}$$

$$N = L/B.$$

## EXAMPLES OF FLAT-SLIDER ANALYSIS

To illustrate briefly the use of the equations, let it be desired to analyze the lubrication of a slider for which $N = 1.0$ and $b = 1.5$. For this slider, $M = -12.0$ and $n$ is thus found to be 3.46. The pressure functions are then

$$f(x_1) = \frac{1.5x_1(x_1 - 1)}{(3.5)(1.5x_1+1)^2}$$

$$f(z_1) = 1 - 0.97(e^{-3.46z_1} + e^{-3.46(1-z_1)}).$$

Graphs of these two functions can be drawn, and by finding curves of constant $f(x_1) \cdot f(z_1)$ the isobars over the slider surface can be located. This has been done in Fig. 2. The numbers on the isobars are values of $f(x_1) \cdot f(z_1)$. The actual pressure at each isobar is, of course, this value of $f(x_1) \cdot f(z_1)$ multiplied by the quantity $A$.

For the case of wide sliders, where $N$ becomes small, the analysis by means of the rigorous equations, or even by the approximate treatment in reference 3 becomes more difficult, but that with the new equations becomes simpler. For a slider where $N = 0.01$ and $b = 1.5$, $M$ is still $-12.0$ but $n$ is now 346. Then,

$$f(z_1) = 1 - (e^{-346z_1} + e^{-346(1-z_1)}).$$

It is observed that the value of this function, although zero at $z_1 = 0$ and 1.0, rises very rapidly to 1 between these endpoints; at $z_1 = 0.02$ or 0.98, for example, the value of $f(z_1)$ is 0.999. The function $f(z_1)$ may therefore be considered as a correction to be applied to the ends of a slider which would otherwise be considered infinite. Under such conditions, the value of $f(z_1)$ should be directly dependent upon $z_1/N$, the actual width represented by $z_1$; that this is actually the case is illustrated by finding the point at which $f(z_1) = 0.999$ for a slider for which $N = 0.0001$ and $b = 1.5$. This value turns out to be $z_1 = 0.0002$, which is 0.01 times that for a slider of $N = 0.01$. Since, however, this second slider is 100 times as wide, the actual distance from the ends of both sliders to the points at which $f(z_1) = 0.999$ is the same. This type of relationship in fact exists up to the point where the quantity

$$(e^n - 1)/(e^n - e^{-n})$$

becomes significantly different from 1.0.

## EXAMPLE OF CURVED-SLIDER ANALYSIS

For purposes of illustration, a slider will be analyzed for which $\beta = (2.50)^{x_1}$ and $N = 0.5$. The

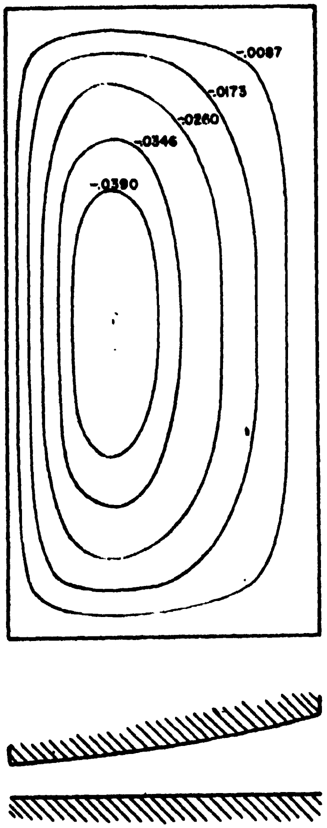


FIG. 3.

customary approximate method leads to a series of values for $f(x_1)$:

| $x_1$ | $\beta$ | $f(x_1)$ |
|---|---|---|
| 0.0 | 1.000 | 0.0000 |
| 0.1 | 1.096 | −0.0273 |
| 0.2 | 1.201 | −0.0414 |
| 0.3 | 1.316 | −0.0466 |
| 0.4 | 1.443 | −0.0459 |
| 0.5 | 1.582 | −0.0414 |
| 0.6 | 1.733 | −0.0349 |
| 0.7 | 1.899 | −0.0273 |
| 0.8 | 2.081 | −0.0192 |
| 0.9 | 2.281 | −0.0112 |
| 1.0 | 2.500 | 0.0000 |

By the use of Lagrange's interpolation formula, $x_m$ is found to be 0.338 and $f(x_m) = -0.0470$. At this point $\beta = 1.363$ and $d\beta/dx_1 = 1.249$

$$M = \frac{1.249}{(-0.0470)(1.363)^3} = -10.50.$$

The value of $n$ is then

$$n = (10.50)^{\frac{1}{2}}/0.5 = 6.482.$$

All the necessary constants have now been calculated. The shapes of the isobars over the surface are illustrated in Fig. 3. The value marked on each isobar gives, when multiplied by $A$, the magnitude of the pressure at that isobar.

TABLE I.

|  | $b=1.0$ |  |  | $b=2.0$ |  |  |
|---|---|---|---|---|---|---|
| $N$ | $W/W_\infty$ (new) | $W/W_\infty$ (actual) | Error % | $W/W_\infty$ (new) | $W/W_\infty$ (actual) | Error % |
| 0 | 1.000 | 1.000 | 0 | 1.000 | 1.000 | 0 |
| 0.5 | 0.687 | 0.695 | −1.2 | 0.735 | 0.714 | 2.9 |
| 1.0 | 0.421 | 0.434 | −3.0 | 0.493 | 0.460 | 7.2 |
| 1.5 | 0.259 | 0.279 | −7.2 | 0.322 | 0.297 | 8.4 |
| 2.0 | 0.168 | 0.185 | −9.2 | 0.219 | 0.208 | 5.3 |

## APPLICABILITY OF EQUATIONS

Although the new equations are especially useful in analyzing cases that are difficult or impossible to solve by the more exact and involved methods, the only available accurate data are for flat sliders, and these must consequently serve as a basis for estimating the accuracy of the solutions obtained by the new method. The comparison thus made is presumed to hold also for curved sliders.

### $W/W_\infty$ Ratio

By $W/W_\infty$ ratio is designated the ratio of the total load carried by the slider to that which would be carried by an equal length of an infinitely wide slider. Since $f(x_1)$ in the new equations is the same as that for an infinite slider, the $W/W_\infty$ ratio is merely the integral of $f(z_1)$ (Eq. (7)), which is

$$\frac{W}{W_\infty} = 1 - \frac{2}{n}\frac{(1-e^{-n})^2}{(1-e^{-2n})}. \qquad (12)$$

Values computed from this equation are compared to those computed by Michell's method for $b=1.0$ and 2.0 in reference 3, page 137 (Table I).

TABLE II.

| $b=2.0$ | $N$ | $D$ (new) | $D$ (actual) | Error % |
|---|---|---|---|---|
|  | 0 | −0.0246 | −0.0246 | 0 |
|  | 0.5 | −0.0181 | −0.0175 | 3.4 |
|  | 1.0 | −0.0122 | −0.0113 | 8.0 |
|  | 1.5 | −0.0079 | −0.0073 | 8.2 |
|  | 2.0 | −0.0054 | −0.0051 | 5.9 |

It will be shown later that the error in the $W/W_\infty$ ratio is of the same magnitude as that in many other derived characteristics of the slider bearing. Figure 4 compares values of $W/W_\infty$ obtained with the new equations to those found by several investigators with various methods, for $N=0.5$, 1.0, and 2.0. The values due to Kingsbury are from "Optimum Conditions for Journal Bearings"[7] and those credited to Muskat, Morgan, and Meres were actually found by a series of graphical calculations with the charts in *Lubrication*,[8] pages 162–66. The latter cannot therefore be considered extremely accurate, but they extend over a greater range than the other published values.

It can be seen that the ratio given by the new equations is in good agreement with those computed by other methods.

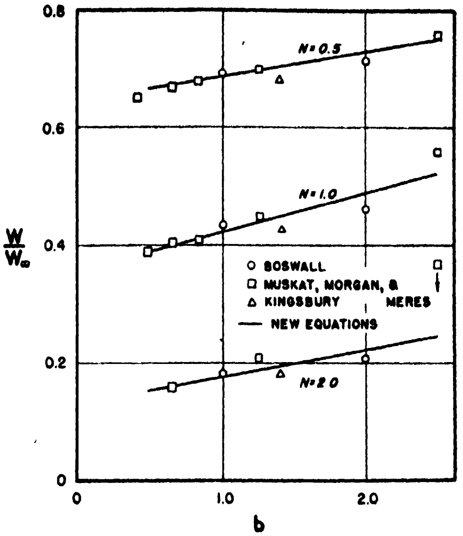


FIG. 4.

### Total Load

The total load carried by the slider is given by the expression

$$W=ALBD,$$

following the notation of Boswall. The new method yields, for the value of $D$,

$$D=\int_0^1 f(z_1)dz_1 \int_0^1 f(x_1)dx_1,$$

because the variables $x_1$ and $z_1$ are independent. Comparison of the values of $D$ from the new and from the Michell solution is sufficient because the factors $A$, $L$, and $B$ are the same for both methods. The accurate data are again obtained from reference 3, page 137 (Table II). It could have been expected that the error in total load be somewhat greater than that in $W/W_\infty$ ratio, because here the variation of $f(x_1)$ from the actual form is also taken into account. Nevertheless, the accuracy is still surprisingly good.

TABLE III.

| $b=2.0$ | $N$ | $\bar{x}_1$ (new) | $\bar{x}_1$ (actual) | Error % |
|---|---|---|---|---|
| | 0 | 0.392 | 0.392 | 0 |
| | 0.5 | 0.392 | 0.386 | 1.5 |
| | 1.0 | 0.392 | 0.374 | 4.9 |
| | 1.5 | 0.392 | 0.359 | 10.5 |
| | 2.0 | 0.392 | 0.341 | 12.6 |

### Center of Pressure ($\bar{x}_1$)

The center of pressure is the value of $x_1$ at which the integrated effect of the pressure is directed. It is thus the line along which the slider would be pivoted. This depends, in the new method, only upon $f(x_1)$ which keeps the same form regardless of $N$. The value of $\bar{x}_1$ is therefore a constant depending only upon $b$, and is equal to $\bar{x}_1$ for an infinite slider. It is, of course, known that the value of $\bar{x}_1$ is actually dependent upon $N$, but the error introduced by assuming it constant remains to be found. The actual values are obtained from the same source previously quoted (Table III). The error in $\bar{x}_1$ is still not excessively great, and the assumption of the constant value would lead to no serious difficulty in most cases.

TABLE IV.

| $b=2.0$ | $N$ | $H$ (new) | $H$ (actual) | Error % |
|---|---|---|---|---|
| | 0 | 1.82 | 1.82 | 0 |
| | 0.5 | 2.00 | 2.02 | −1.0 |
| | 1.0 | 2.30 | 2.38 | −3.4 |
| | 1.5 | 2.74 | 2.83 | −3.2 |
| | 2.0 | 3.23 | 3.32 | −2.7 |

---

[7] Albert Kingsbury, "Optimum conditions for journal bearings," Trans. ASME RP-54-7, **54**, 123–148 (1932).

[8] A. E. Norton, *Lubrication* (McGraw-Hill Book Company, Inc. New York and London, 1942), pp. 162–166.

TABLE V.

| $b=2.0$ | $N$ | $V_0$ (new) | $V_0$ (actual) | Error % |
|---|---|---|---|---|
| | 0 | 0.288 | 0.288 | 0 |
| | 0.5 | 0.225 | 0.227 | −0.9 |
| | 1.0 | 0.169 | 0.172 | −1.8 |
| | 1.5 | 0.126 | 0.130 | −3.1 |
| | 2.0 | 0.100 | 0.115 | −14.0 |
| $b=2.0$ | $N$ | $V_i$ (new) | $V_i$ (actual) | Error % |
| | 0 | 0.288 | 0.288 | 0 |
| | 0.5 | 0.313 | 0.313 | 0 |
| | 1.0 | 0.306 | 0.299 | 2.3 |
| | 1.5 | 0.274 | 0.262 | 4.7 |
| | 2.0 | 0.240 | 0.245 | 2.0 |

## Frictional Force

In reference 3, page 145, the frictional resistance is given by the equation

$$R=W\left(\frac{\mu_0 U}{W_1}\right)^{\frac{1}{2}}H,$$

where $H$ is an expression involving integrals of $f(x_1)$ and $f(z_1)$. As in the case of total load, a comparison of values of $H$ is sufficient, since the remaining terms are identical for both methods (Table IV). Here, the error is very small indeed.

## Quantity of Lubricant

The following expressions are given for the quantity of lubricant in reference 3:

$$V_1=\tfrac{1}{2}(6D)^{\frac{1}{2}}\left[\beta_1-\frac{\beta_1^3}{\epsilon}G_1\int_0^1 f(z_1)dz_1\right]$$

where

$$G_1=\left[\frac{df(x_1)}{dx_1}\right]_{x_1=1},$$

$$V_0=\tfrac{1}{2}(6D)^{\frac{1}{2}}\left[1-G_0\int_0^1 f(z_1)dz_1\right]$$

where

$$G_0=\left[\frac{df(x_1)}{dx_1}\right]_{x_1=0}.$$

$V_1$ and $V_0$ are the inlet and outlet quantities, respectively. Substitution of the proper values into these equations yields the results given in Table V. The error in $V_0$ is too large at $N=2.0$ but is certainly sufficiently small at $N=1.5$. The error in $V_1$ is small enough over the entire considered range of $N$.

TABLE VI.

| $b=2.0$ | $N$ | $W/W_\infty$ | $D$ | $\bar{x}_1$ | $H$ | $V_i$ | $V_0$ |
|---|---|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0.5 | 2.9 | 3.4 | 1.5 | −1.0 | 0 | −0.9 |
| | 1.0 | 7.2 | 8.0 | 4.9 | −3.4 | 2.3 | −1.8 |
| | 1.5 | 8.4 | 8.2 | 10.5 | −3.2 | 4.7 | −3.1 |
| | 2.0 | 5.3 | 5.9 | 12.6 | −2.7 | 2.0 | −14.0 |

## Estimation of Error

It was noted previously that the value of $W/W_\infty$ is very easily calculated by the new method, and it would be desirable to use the error in this quantity as a criterion of the accuracy of the method. For purposes of comparison, then, the errors are tabulated in Table VI. It is considered that the $W/W_\infty$ ratio forms a fair, if not slightly pessimistic, criterion of the accuracy of $D$, $H$, and $V_1$, while for $\bar{x}_1$ and $V_0$ the error is likely to be greater in magnitude than that estimated from $W/W_\infty$ when $N$ becomes large. For the great majority of sliders, for which $N\leqslant 1.0$, the $W/W_\infty$ criterion is satisfactory.

Since Fig. 4 shows good agreement between actual and computed $W/W_\infty$ ratios even at $N=2.0$, it can be concluded that the new equations hold with good accuracy for all sliders for which $N$ is 1.0 or less, and they provide an excellent basis for estimating the behavior of even narrower sliders.

## ACKNOWLEDGMENTS

The original work on this project was carried out at the National Advisory Committee for Aeronautics, Aircraft Engine Research Laboratory, Cleveland, Ohio, and was published as ARR E4122. The present report constitutes a refinement in treatment and a considerable expansion of that work.

The assistance of Miss Thelma Bresser in reading the proofs is gratefully acknowledged.

# Letters to the Editor

## Measurements of Base Pressure on a Missile in Free Flight*

F. K. Hill and R. A. Alpher
*Applied Physics Laboratory, The Johns Hopkins University, Silver Spring, Maryland*
February 24, 1947

In a series of experiments now under way, missiles of simple geometry (see Fig. 1), differing only in length, are rocket-launched to supersonic velocities; telemetering installed in the missiles transmits to a ground receiver the aerodynamic data measured during flight. Among the significant data being obtained are base pressures; such data are important in the extrapolation of wind-tunnel studies on models to full-scale performance of missiles, where large Reynolds numbers are encountered under ordinary atmospheric conditions.

The missiles have a 14-caliber ogival nose joining tangentially to a cylindrical body of 4-inch diameter, the base of the cylinder being flat and closed. Four fins of bi-convex section are added to give flight stability. The yaw of the missile (in the plane including the missile axis and base pressure measurement port) with respect to the flight path is determined by differential pressure measurements near the nose tip. These differential pressure measurements may be correlated with yaw by means of laboratory experiments in an open supersonic jet or in wind tunnels, as well as by calculations based on linearized supersonic flow theory. It is found in these experiments that the yaw angles encountered are of the order of 1° or less and that no measurable effects caused by angle of attack are observed on base pressure data here reported. Missile velocities are obtained both by an optical tracking system and by stagnation pressure measurement on the missile with a Pitot tube.

Base pressures were measured at a 0.040″ port located ⅜″ from the center line of the missile. Observed base pressures were converted to pressure coefficients by dividing them by the dynamic head (namely, $\frac{1}{2}\times$density$\times$velocity$^2$); corrections caused by change in altitude during flight were made. Coefficients thus computed are shown (Fig. 2) for missiles of the first two lengths of the series to be studied. It is to be noted that the coefficients show at most a small variation with Mach number in the supersonic range covered by the tests, where the value is about 0.3. This result is significantly different from previously used theoretical values.[1] While present information indicates that pressure is essentially independent of position on the base of such missiles, an experimental check of this will be made in free flight.

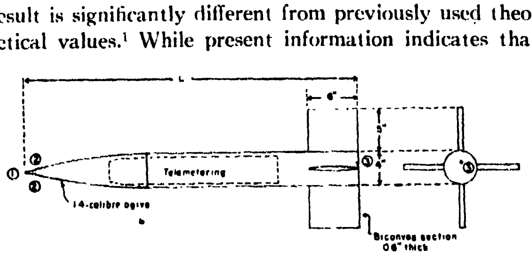


Fig. 1. Scale effect test missile. Telemetering to record: (1) Pitot tube stagnation pressure, (2) differential pressure at nose ports, (3) base pressure at port ⅜″ from center. All pressure ports in the same plane through missile axis.

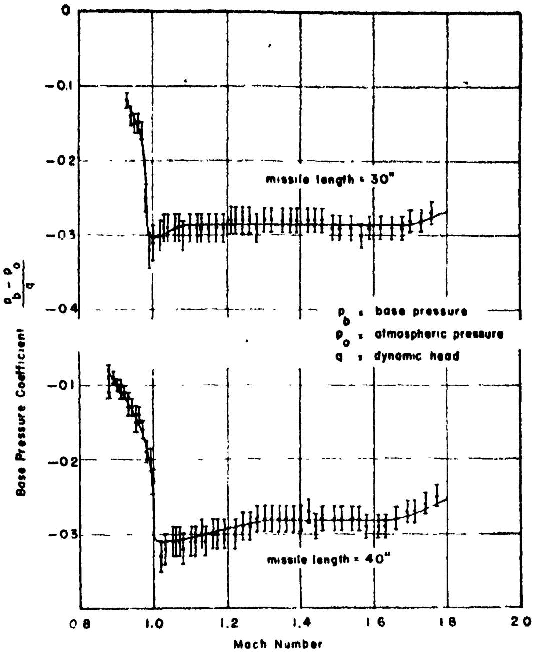


Fig. 2. The probable errors indicated are based upon fairly well-known inaccuracies in the determination of the dynamic head and static pressure throughout flight, and an assumed inaccuracy in the telemetering records based on laboratory tests. All errors are believed to be random.

Similar missiles with lengths up to 120″, as well as a series differing only in diameter, and a series of homologous models, are to be studied, with maximum velocities corresponding to Mach number 3.

This study is part of a general investigation of scale effects which is to be reported in detail after completion of all the tests. It provides the first direct measure of base pressure on such high speed missiles.

---

* Work being performed by a group under contract NOrd 7386 with the Bureau of Ordnance, U. S. Navy.

[1] 0.2 given by von Kármán and Moore, Trans. ASME 54, 303 (1932).

---

## Direction of Maximum Crystal Elongation during Metal Cutting

G. H. Townend
*Metropolitan-Vickers Electrical Company, Ltd., Manchester, England*
March 17, 1947

In an interesting paper by Dr. M. E. Merchant on "Mechanics of the Metal Cutting Process,"[1] an error was made in the expression connecting shear angle, rake angle, and direction of maximum crystal elongation (Eq. (1), page 269). I have written to Dr. Merchant and he confirms the new expression derived below.

Let $\alpha$ = rake angle,
$\phi$ = the angle between the shear plane and the surface of the workpiece.

Consider a section of the workpiece $ABCD$ such that $\angle CBD = 90°$. Then after shearing this section will become the parallelogram $BFEC$. The dimension perpendicular to the shear plane must be unchanged during shearing.

$$\therefore \quad BD = BH = a.$$

Consider a line $CG$ in the chip making an angle $\psi$ with the plane of shear. Then the corresponding line in the workpiece is $DJ$. The condition that $CG$ shall be the direction of maximum crystal elongation is that $CG/DJ$ is a maximum.

$$CG = a \operatorname{cosec}\psi,$$
$$DJ = \{a^2 + (a \cot\phi - a \cot\psi + a \tan(\phi - \alpha)^2\}^{\frac{1}{2}}.$$

The condition is that

$$\{\sin^2\psi + (\cot\phi \sin\psi - \cos\psi + \tan(\phi - \alpha) \sin\psi)^2\}^{-\frac{1}{2}}$$

is a maximum. That is,

$$\sin^2\psi + \{\sin\psi(\cot\phi + \tan(\phi - \alpha) - \cos\psi\}^2$$

is a minimum. Let

$$\cot\phi + \tan(\phi - \alpha) = \epsilon,$$
$$\epsilon^2 \sin 2\psi - 2\epsilon \cos 2\psi = 0.$$

If $\epsilon \neq 0$

$$2 \cot 2\psi = \epsilon,$$
$$2 \cot 2\psi = \cot\phi + \tan(\phi - \alpha).$$

Alternatively it is easily shown that a circle on the workpiece becomes an ellipse on the chip and that the principal axes of the ellipse make an angle $\psi$ with the shear plane, where

$$2 \cot 2\psi = \cot\phi + \tan(\phi - \alpha).$$

---

[1] J. App. Phys. 16, 267 (1945).

---

# Here and There

### Printing Electronic Circuits

According to a recent release from the National Bureau of Standards prepared by Cledo Brunetti and Philip J. Franklin, new techniques, developed by the National Bureau of Standards in cooperation with electronic manufacturers, for electronic devices promise to simplify production methods and to accelerate the manufacture of cheaper, more compact, and more uniform electronic equipment. These techniques, representing the perfection of a method of printing wiring and circuit components on an insulated surface, is one of the new processes and techniques evolved from research and development during the war.

The methods are applicable in the design of numerous industrial and commercial devices where extreme ruggedness and small size are imperative. By means of the printing process the electronic circuit of a conventional radio—now characterized by a maze of wires, resistors, inductors, and condensers—may be replaced by a compact circuit printed on a small flat plate, eliminating much of the third dimension and making it simple to check and repair. Other specific applications lie in the fields of pocket radios, personal radio telephones, miniature hearing aids, meteorological instruments, and various miniature electronic control circuits.

Present investigations at the National Bureau of Standards of the various means available for reducing the size of electronic assemblies and for mechanization of chassis wiring reveal a number of feasible and practical electronic printing methods. Variations of the stenciled screen-printing process include hand printing or spraying of the components and leads on suitable surfaces. Another method involves the use of photography and an abrasive blast. A third is an adaptation of the chemical methods of silvering mirrors. Mechanical processes include metal spraying, metal sputtering or evaporation, the use of foils, and stamping. As an example, loop antennas are now stamped out in a single operation by a variation of the printed wiring techniques and have shown improved performance over wire-wound loops in addition to reduced cost of production. The principal effect of these methods is to reduce electronic circuit wiring to two dimensions. The effect is enhanced where it is possible to employ sub-miniature tubes and compact associated components.

### New Appointments

Leonard I. Schiff, theoretical physicist who worked on the Los Alamos atom bomb project, has joined the faculty of Stanford University as associate professor of physics.

H. B. G. Casimir, co-director of the Philips Research Laboratories, Eindhoven, Holland, is delivering a series of invitation lectures at Johns Hopkins University on the properties of matter at low temperatures and on problems in quantum electrodynamics.

Leo L. Beranek, formerly director of the electro-acoustic and systems research laboratories of Harvard University, has been appointed associate professor of communications engineering in the Department of Electrical Engineering of the Massachusetts Institute of Technology.

Alfred J. Reis, Austrian-born scientist, was recently appointed a research specialist in the Engineering Experiment Station at Rutgers University. He will conduct a series of studies in the engineering properties of materials and in powder metallurgy.

John A. Hipple has been named chief of the Atomic Physics Section at the National Bureau of Standards.

John L. Abbott recently joined North American Philips Company, Inc., as application engineer in the industrial x-ray division.

### Awards

James I. Hoffman of the National Bureau of Standards has been awarded the 1946 Hillebrand Prize for significant contributions to chemical science.

The John Price Wetherill and the Howard N. Potts medals of The Franklin Institute of Philadelphia were awarded at ceremonies on April 16. Kenneth S. M. Davidson of Stevens Institute of Technology received the Wetherill medal for his contributions to the improved design of ships through experiments on ship hull models. Vladimir K. Zworykin, director of electronic research of the Radio Corporation of America, and Robert H. Kent, associate director of the ballistics research laboratories, Aberdeen Proving Ground, were each awarded a Potts medal. Zworykin received the medal for his inventions of the iconoscope and kinescope, which are essential to modern commercial television. Kent was honored for contributions to the science of ballistics and particularly for his effective operation of the Aberdeen Wind Tunnel, which is doing a large part of the supersonic testing for the armed forces.

The seven winners of Frank B. Jewett fellowships for research in the physical sciences for 1947–1948, announced by the American Telephone and Telegraph Company, are as follows: M. G. Ettlinger of California Institute of Technology and Austin, Texas; Wallace D. Hayes of California Institute of Technology and Palo Alto, California; Paul Olum of Harvard University and Winchester, Massachusetts; Aadne Ore of Yale University and Oslo, Norway; Alfred Schild of the Carnegie Institute of Technology and Pittsburgh, Pennsylvania; Robert L. Scott of the University of California and Santa Ana, California; and Edwin H. Spanier of the University of Michigan and St. Paul, Minnesota. Two of the winners are chemists, two physicists, two mathematicians, and the seventh is an aeronautical engineer.

### General Electric-Yale University Joint Research

At the Dunham Laboratory of Electrical Engineering at Yale University, research is being conducted on controls for a flight simulator or dummy rocket for duplicating flight characteristics of a rocket missile, as part of General Electric Company's long-range rocket research program. The project at Yale is under the direction of John L. Bower, Assistant Professor of Electrical Engineering.

### Biological Photographic Exhibition

An international exhibition of biological photography of interest to all photographers in the scientific field will be held September 10–27, 1947, under the sponsorship of the Biological Photographic Association at the Rochester Public Library, Rochester, New York. Three judges who are prominent in their fields will select all entries. Membership in the Association is not a requisite for participation in the exhibition. Entry blanks for contributors may be obtained from Mr. H. Lou Gibson, 343 State Street, Rochester 4, New York. The entry fee is $1, and the last day for receiving entries is August 1, 1947.

Information concerning the convention program of the Biological Photographic Association to be held September 10–13 may be obtained from Mr. John V. Butterfield, 635 St. Paul Street, Rochester 4, New York.

### General Electric Educational Fund Awards

Research fellowships totaling $20,500 have been awarded to fourteen graduate students by the General Electric Educational Fund. Eight of the winners received Charles A. Coffin Fellowships awarded for research in the fields of electricity, physics, and physical chemistry. The other six young men obtained Gerard Swope Fellowships for advanced study in industrial management, engineering, the physical sciences, and other scientific and industrial fields.

### M.I.T. Summer Session in Applied Mathematics

Massachusetts Institute of Technology has announced a special summer session in applied mathematics from August 4 to September 19, 1947. The following graduate courses will be offered:

1. **Theory of Plates and Shells.** By Professors F. B. Hildebrand and E. Reissner.
2. **Theoretical Hydromechanics.** By Professors E. Reissner and C. C. Lin.
3. **Advanced Topics in Applied Mathematics.** By Professors J. L. Synge, H. Reissner, W. Prager, F. B. Hildebrand and P. D. Crout.
4. **Tensors in Mechanics.** By Professor D. J. Struik.

Further information concerning the program may be obtained by writing to Department of Mathematics, Massachusetts Institute of Technology, Cambridge 39, Massachusetts.

### Centenary of Association of Engineers

The Association of Engineers, with headquarters at the University of Liége in Belgium, celebrates this year its 100th anniversary. For the occasion it is sponsoring a congress and an exhibition to which engineers and industry representatives are cordially invited. The congress will take place in Liége from August 30 to September 13, and is entitled "Past, Present and Future of Our Industry." For further information address Association des Ingenieurs, 12, Quai Paul Van Hoegaerden, Liége, Belgium.

### National Electronics Conference

The National Electronics Conference, Inc., whose purpose is to serve as a national forum on electronic developments and their application, is sponsored jointly by Illinois Institute of Technology, Northwestern University, American Institute of Electrical Engineers, Institute of Radio Engineers, and the University of Illinois, with the Chicago Technical Societies Council a cooperating organization. Plans are now being formulated for the 1947 National Electronics Conference, which will be held on November 3, 4, and 5 at the Edgewater Beach Hotel, Chicago.

### Course in Statistical Quality Control

An advanced eight-day course in quality control by statistical methods is to be presented June 27 through July 5 at Purdue University, Lafayette, Indiana. It is intended for those who already have some acquaintance with statistical quality control from such a course as the War Production Board organized during the war. Anyone desiring further information should write to Professor I. W. Burr at Purdue.

### Marconi Anniversary

An International Congress will celebrate the fiftieth anniversary of the discovery of radio by Marconi, September 28 to October 5, 1947, in Rome, Italy. The event is being organized by the Italian National Council of Research. It aims at providing a complete picture of the progress of radio in the world from the scientific, technical, and industrial point of view, and giving world radio experts the opportunity of meeting and studying plans for a future international collaboration.

### The Bulletin of Mathematical Biophysics

Following is the table of contents for *The Bulletin of Mathematical Biophysics* for June 1947:

**Mathematical Theory of Motivation Interactions of Two Individuals: II.** ANATOL RAPOPORT

**A Theory of Membrane Permeability: III. The Effect of Hydrostatic Pressure.** INGRAM BLOCH

**The Mechanism of the Middle Ear: II. The Drum.** MARTINUS H. M. ESSER

**A Mathematical Description of Metabolizing Systems: II.** HERMAN BRANSON

**Outline of a Matrix Calculus for Neural Nets: II.** H. D. LANDAHL

The University of Chicago Press, Chicago, Illinois, Volume 9, Number 2.

### Meeting of the Division of Solid State Physics

A Division of Solid State Physics in the American Physical Society, recently authorized by the Council, is in the process of formation and a corresponding announcement is being sent out to the members of the society. The organizing committee of the division is planning for a meeting of the division in conjunction with the Montreal meeting of the American Physical Society June 19 to 21, 1947. In view of the fact that the American Society for X-Ray and Electron Diffraction is having a meeting near Montreal starting Monday, June 23, the meeting of the division will take place on Saturday, June 21, so as to make it convenient for those interested in solid state to attend both meetings.

The program will be composed of four invited papers:

"Diffraction of Neutrons in Crystals," F. SEITZ.

"Theory of Crystal Rectifiers with Application to Copper Oxide," J. BARDEEN.

"The Relaxation Spectrum of Metals," C. ZENER.

"Nuclear Magnetic Resonance Absorption in Solids and Liquids," E. M. PURCELL,

and contributed papers in the field of solids. It is suggested that those who have papers in the field of the division would present them at the June meeting in Montreal.

The Organizing Committee of the Division of Solid State:

| | |
|---|---|
| T. A. Read | S. Siegel |
| F. Seitz | R. Smoluchowski |
| W. Shockley | C. Zener |

### Calendar of Meetings

**May**

| | |
|---|---|
| 26–27 | Institute of Aeronautical Sciences, Detroit, Michigan |
| 31 | American Physical Society, New England Section, Amherst, Massachusetts |

**June**

| | |
|---|---|
| 1–6 | Society of Automotive Engineers, French Lick, Indiana |
| 8–10 | American Society of Refrigerating Engineers, Los Angeles, California |
| 9–13 | American Institute of Electrical Engineers, Montreal, Quebec, Canada |
| 9–14 | Symposium on Molecular Structure and Microscopy, Ohio State University, Columbus, Ohio |
| 12–14 | Canadian Association of Professional Physicists, University of Western Ontario, London, Canada |
| 16–19 | American Society of Mechanical Engineers, Chicago, Illinois |
| 16–20 | American Society for Testing Materials, Atlantic City, New Jersey |
| 18–20 | American Association of Physics Teachers, Regional Meeting, University of Minnesota, Minneapolis, Minnesota |
| 18–21 | Society for the Promotion of Engineering Education, Minneapolis, Minnesota |
| 23–25 | American Society for X-Ray and Electron Diffraction, Montreal, Quebec, Canada |

**July**

| | |
|---|---|
| 11–12 | American Physical Society, Stanford University, California |
| 16–19 | American Society of Civil Engineers, Duluth, Minnesota |

**August**

| | |
|---|---|
| 21–23 | Society of Automotive Engineers, Los Angeles, California (Transportation and Maintenance Meeting) |
| 26–29 | American Institute of Electrical Engineers, San Diego, California |

## New Booklets

Dow Corning Corporation, Midland, Michigan, has announced a new technical pamphlet, *DC Antifoam A*, which discusses a silicone compound recently developed for killing foam in aqueous systems. 4 pages.

The December 1946 issue of *The Frontier*, published by the Armour Research Foundation of Illinois Institute of Technology, Chicago 16, Illinois, includes the following articles: "A Unique Torquemeter," "Classification of Analytical Methods," "Processing and Use of Zircon for Refractory Electrical Insulating Compositions." 20 pages.

*Nickel Steel Topics* for February 1947, published by The International Nickel Company, Inc., 67 Wall Street, New York 5, New York, features an article on the "refrigerator car of the future," built largely of stainless and low nickel alloy steels. 12 pages.

Fish-Schurman Corporation, 230 East 45th Street, New York 17, New York, makers of raw optical, technical, and scientific glass and optical parts and specialties, recently issued bulletin MI-318 on *FS Multi-Layer Interference Films; Dichroic and Achromatic Beam Splitters*. 2 pages.

*The Aerovox Research Worker*, monthly house organ of the Aerovox Corporation, New Bedford, Massachusetts, in its December 1946 edition discusses "The Rectifier-Amplifier V. T. Voltmeter." 4 pages.

**Pittsburgh Plate Glass Company, Columbia Chemical Division, Fifth Avenue at Bellefield, Pittsburgh 13, Pennsylvania, has published *Form A-100*, a new Columbia Caustic Soda Book, 72 pages, printed in three colors. Many photographs. Write for a copy on your company letterhead.**

Journal of
Applied Physics

Volume 18, Number 6 June, 1947

# Crystal Pattern Synthesis by an Approximate Summation of Fourier Series

JOSEPH S. LUKESH
*Crystallographic Laboratory, Department of Geology, Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Received February 10, 1947)

A faithful representation of the locations of the peaks of electron density maps can be obtained rapidly by determining for each point in the projected unit cell the algebraic sum of the amplitudes of those reflections which are contributing either a maximum or a minimum at each point. Since maxima and minima do not, in general, fall at the specific points of the grid, an approximation must be introduced. Using the semi-polar form of Fourier series, the approximate summation can be accomplished with the use of one hundred and fifty strips. The method was tested by determining the peak locations in the structure of diopside projected on 001. Potential applications of the approximate summation include rapid checking of the validity of assumed crystal structures.

## INTRODUCTION

CRYSTAL patterns can be synthesized conveniently by determining the projected electron density through the summation of a Fourier series of the following general type:

$$\rho(x, y) = (1/A)\sum_h \sum_k F_{(hk0)} \cos(2\pi/N)(hx+ky), \tag{1}$$

which gives the projection normal to the *c* axis. Other axes are treated by appropriate choice of the indices, *h*, *k*, or *l*. Equation (1) is valid only when the projection is centrosymmetrical. The non-centrosymmetrical case will not be considered here; however, it can be treated in a similar manner by applying the same reasoning to the sine terms. The form of the contribution of a given reflection, *hk0*, to the total electron density is invariant, regardless of the actual crystal structure. It consists of a sinusoidal alternation of positive and negative density. Actually it is a cosine surface whose periodicity in each axial direction is given by the appropriate index. The different distributions of electron density in different crystals arise from varying amplitudes of the cosine surfaces and from different arrays of contributing reflections. Obviously, in order that an accumulation of electron density occur at any point in the projected unit cell, the sum of the positive solutions of the terms of Eq. (1) must be greater than the sum of the negative solutions. In order that a high peak occur, not only must a large number of reflections give positive solutions, but also they must be near their maxima. Similarly, in order that there be little or no accumulation of electron density, either most reflections are contributing very little, or those near their maxima are compensated for by others near their minima. This suggests that if one were to list for each point in the unit cell the amplitudes of all reflections that are contributing a maximum and all that are near a minimum, one might obtain a faithful representation of the electron density distribution merely by subtracting the negative terms from the positive. If this hypothesis is valid, the labor of computation is reduced to the listing and totalling of the appropriate amplitudes. Of course, detail will be lost in the regions of low electron density, but these regions are unimportant in at least the early stages of crystal structure analysis. In many cases, location of the strongest peaks will be of material aid in the solution of a crystal structure.

## THEORY

In an earlier paper the writer[1] developed a semi-polar form of the series of Eq. (1). This concept, while not essential to the approximate method to be discussed, offers a number of advantages in computational simplicity and will be followed in this report. If the point, $x, y$, is specified by the coordinate, $x$, and the slope, $c$, of the line from the origin to the point, Eq. (1) may be written:

$$\rho(x, c)=(1/A)\sum_h \sum_k F_{(hk0)} \cos(2\pi/N)x(h+ck). \tag{2}$$

The half-unit cell for which the summation is to be made is divided into four segments, and the form of the cosine term varies for each. For a given plane and a given radial line, Eq. (2) reduces to

$$\rho(x, c)=(1/A)\sum_h \sum_k F_{(hk0)} \cos(2\pi/N)Cx, \tag{3}$$

where $C$ is a constant and $x$ is an integer. In the article cited[1] it was pointed out that the conventional rectilinear grid is inconvenient, since the slopes of the lines from the origin to the various points are not suited for easy calculation. Furthermore, only a few points occur along any one line. For these reasons, a semi-polar grid was described in which the points at which Eq. (3) is to be evaluated are defined by the intersections of radial lines and either the $x$ or $y$ coordinates. If the axes are divided into thirty points (a sufficient number since the projection is only an approximation), the number, $N$, in Eq. (3) is thirty. For this spacing of points, it is convenient to choose radii whose slopes are .1, .2, .3, etc., and the constant, $C$, can be determined from the indices of a plane merely by inspection. Thus, for the reflection, 320, the constant along the radius whose slope is .4 is equal to 3 plus .4(2), or 3.8.

If Eq. (3) is plotted as a function of $x$ for a given amplitude, $F$, and a given constant, $C$, a cosine curve results. This is the form of the density contribution of a reflection, $hk0$, along a radial line of slope, $c$. The amplitude is determined by $F$ and the periodicity by $C$. It is obvious that maxima and minima occur at various points along the radius. Since the positions of the maxima and minima depend only on the indices of the reflection and on the slope of the radius, they can be determined in advance and tabulated for use in all structure determinations. Because, as was pointed out in the article previously cited,[1] there are only one hundred and fifty independent values of $C$ for the grid chosen, the table will not be extensive.

In general, the maxima and minima will not occur at integral values of the coordinate, $x$. This necessitates the introduction of an approximation, since it is at the integral values of $x$ that the summation is being made. For the purposes of this report, it has been assumed, quite arbitrarily, that a plane contributes to the electron density only at those points where its amplitude times the appropriate value of the cosine is fifty or more percent of the maximum. Similarly, a plane subtracts from the density only when the product is fifty or more percent of the minimum. In other words, a reflection is assumed to be ineffectual when the product of its amplitude times the cosine is between zero and either plus or minus one-half of $F$. Furthermore, it is assumed that a plane contributes or subtracts the entire value of $F$ when it is effective. This latter assumption eliminates all calculation, and the summation consists of tabulation of the amplitudes of the effective planes at each point and their summation.

## EXAMPLE OF THE APPROXIMATION

In order to test the hypothesis developed above, computations have been made of the electron density of diopside projected on (001). Data for the amplitudes of the various reflections and a map of the conventional summation have been reported by Bragg.[2] Because of the high symmetry of diopside in this projection, only one-eighth of the cell need be considered. However, in order to test the approximation, computations were made for one-quarter of the cell.

---

[1] Joseph S. Lukesh, J. App. Phys. **18**, 321 (1947).

[2] W. L. Bragg, Zeits. f. Krist. **70**, 475 (1929).

TABLE I. Coordinates, $x$, for which the function: $\cos(\pi/15)Cx$ has maxima and minima.

| $C$ | Maxima | Minima | $C$ |
|---|---|---|---|
| 0.0 | 0–15 | None | 30.0 |
| 0.1 | 0–15 | None | 29.9 |
| 0.2 | 0–15 | None | 29.8 |
| 0.3 | 0–15 | None | 29.7 |
| 0.4 | 0–12 | None | 29.6 |
| 0.5 | 0–10 | None | 29.5 |
| 0.6 | 0–8 | None | 29.4 |
| 0.7 | 0–7 | 15 | 29.3 |
| 0.8 | 0–6 | 13–15 | 29.2 |
| 0.9 | 0–5 | 12–15 | 29.1 |
| 1.0 | 0–5 | 10–15 | 29.0 |
| 1.1 | 0–4 | 10–15 | 28.9 |
| 1.2 | 0–4 | 9–15 | 28.8 |
| 1.3 | 0–3 | 8–15 | 28.7 |
| 1.4 | 0–3 | 8–14 | 28.6 |
| 1.5 | 0–3 | 7–13 | 28.5 |
| 1.6 | 0–3 | 7–12 | 28.4 |
| 1.7 | 0–2, 15 | 6–11 | 28.3 |
| 1.8 | 0–2, 14, 15 | 6–11 | 28.2 |
| 1.9 | 0–2, 14, 15 | 6–10 | 28.1 |
| 2.0 | 0–2, 13–15 | 5–10 | 28.0 |
| 2.1 | 0–2, 12–15 | 5–9 | 27.9 |
| 2.2 | 0–2, 12–15 | 5–9 | 27.8 |
| 2.3 | 0–2, 11–15 | 5–8 | 27.7 |
| 2.4 | 0–2, 11–14 | 5–8 | 27.6 |
| 2.5 | 0–2, 10–14 | 4–8 | 27.5 |
| 2.6 | 0, 1, 10–13 | 4–7 | 27.4 |
| 2.7 | 0, 1, 10–12 | 4–7, 15 | 27.3 |
| 2.8 | 0, 1, 9–12 | 4–7, 15 | 27.2 |
| 2.9 | 0, 1, 9–12 | 4–6, 14, 15 | 27.1 |
| 3.0 | 0, 1, 9–11 | 4–6, 14, 15 | 27.0 |
| 3.1 | 0, 1, 9–11 | 4–6, 13–15 | 26.9 |
| 3.2 | 0, 1, 8–10 | 4–6, 13–15 | 26.8 |
| 3.3 | 0, 1, 8–10 | 4–6, 13–15 | 26.7 |
| 3.4 | 0, 1, 8–10 | 3–5, 12–14 | 26.6 |
| 3.5 | 0, 1, 8–10 | 3–5, 12–14 | 26.5 |
| 3.6 | 0, 1, 7–9 | 3–5, 12, 13 | 26.4 |
| 3.7 | 0, 1, 7–9, 15 | 3–5, 11–13 | 26.3 |
| 3.8 | 0, 1, 7–9, 15 | 3–5, 11–13 | 26.2 |
| 3.9 | 0, 1, 7, 8, 15 | 3–5, 11, 12 | 26.1 |
| 4.0 | 0, 1, 7, 8, 14, 15 | 3–5, 10–12 | 26.0 |
| 4.1 | 0, 1, 7, 8, 14, 15 | 3, 4, 10–12 | 25.9 |
| 4.2 | 0, 1, 6–8, 14, 15 | 3, 4, 10, 11 | 25.8 |
| 4.3 | 0, 1, 6–8, 13–15 | 3, 4, 10, 11 | 25.7 |
| 4.4 | 0, 1, 6, 7, 13, 14 | 3, 4, 10, 11 | 25.6 |
| 4.5 | 0, 1, 6, 7, 13, 14 | 3, 4, 9–11 | 25.5 |
| 4.6 | 0, 1, 6, 7, 12–14 | 3, 4, 9, 10 | 25.4 |
| 4.7 | 0, 1, 6, 7, 12, 13 | 3, 4, 9, 10, 15 | 25.3 |
| 4.8 | 0, 1, 6, 7, 12, 13 | 3, 4, 9, 10, 15 | 25.2 |
| 4.9 | 0, 1, 6, 7, 12, 13 | 3, 4, 9, 10, 15 | 25.1 |
| 5.0 | 0, 1, 5–7, 11–13 | 2–4, 8–10, 14, 15 | 25.0 |
| 5.1 | 0, 5, 6, 11, 12 | 2, 3, 8, 9, 14, 15 | 25.9 |
| 5.2 | 0, 5, 6, 11, 12 | 2, 3, 8, 9, 14, 15 | 24.8 |
| 5.3 | 0, 5, 6, 11, 12 | 2, 3, 8, 9, 14, 15 | 24.7 |
| 5.4 | 0, 5, 6, 11, 12 | 2, 3, 8, 9, 13, 14 | 24.6 |
| 5.5 | 0, 5, 6, 10, 11 | 2, 3, 8, 9, 13, 14 | 24.5 |
| 5.6 | 0, 5, 6, 10, 11 | 2, 3, 8, 13, 14 | 24.4 |
| 5.7 | 0, 5, 6, 10, 11, 15 | 2, 3, 8, 13, 14 | 24.3 |
| 5.8 | 0, 5, 6, 10, 11, 15 | 2, 3, 7, 8, 13 | 24.2 |
| 5.9 | 0, 5, 10, 11, 15 | 2, 3, 7, 8, 12, 13 | 24.1 |
| 6.0 | 0, 5, 10, 15 | 2, 3, 7, 8, 12, 13 | 24.0 |
| 6.1 | 0, 5, 10, 14, 15 | 2, 3, 7, 8, 12, 13 | 23.9 |
| 6.2 | 0, 5, 9, 10, 14, 15 | 2, 3, 7, 8, 12 | 23.8 |
| 6.3 | 0, 4, 5, 9, 10, 14, 15 | 2, 3, 7, 12 | 23.7 |
| 6.4 | 0, 4, 5, 9, 10, 14 | 2, 3, 7, 11, 12 | 23.6 |
| 6.5 | 0, 4, 5, 9, 10, 14 | 2, 3, 7, 11, 12 | 23.5 |
| 6.6 | 0, 4, 5, 9, 13, 14 | 2, 3, 7, 11, 12 | 23.4 |
| 6.7 | 0, 4, 5, 9, 13, 14 | 2, 6, 7, 11, 15 | 23.3 |
| 6.8 | 0, 4, 5, 9, 13 | 2, 6, 7, 11, 15 | 23.2 |
| 6.9 | 0, 4, 5, 8, 9, 13 | 2, 6, 7, 11, 15 | 23.1 |
| 7.0 | 0, 4, 5, 8, 9, 13 | 2, 6, 7, 10, 11, 15 | 23.0 |
| 7.1 | 0, 4, 8, 9, 12, 13 | 2, 6, 7, 10, 11, 15 | 22.9 |
| 7.2 | 0, 4, 8, 9, 12, 13 | 2, 6, 10, 11, 14, 15 | 22.8 |
| 7.3 | 0, 4, 8, 12, 13 | 2, 6, 10, 14, 15 | 22.7 |
| 7.4 | 0, 4, 8, 12 | 2, 6, 10, 14 | 22.6 |
| 7.5 | 0, 4, 8, 12 | 2, 6, 10, 14 | 22.5 |
| 7.6 | 0, 4, 8, 12 | 2, 6, 10, 14 | 22.4 |
| 7.7 | 0, 4, 8, 12, 15 | 2, 6, 10, 13, 14 | 22.3 |
| 7.8 | 0, 4, 8, 11, 12, 15 | 2, 6, 9, 10, 13, 14 | 22.2 |
| 7.9 | 0, 4, 7, 8, 11, 12, 15 | 2, 6, 9, 10, 13 | 22.1 |
| 8.0 | 0, 4, 7, 8, 11, 15 | 2, 5, 6, 9, 10, 13 | 22.0 |
| 8.1 | 0, 4, 7, 8, 11, 15 | 2, 5, 6, 9, 13 | 21.9 |
| 8.2 | 0, 4, 7, 11, 15 | 2, 5, 6, 9, 13 | 21.8 |
| 8.3 | 0, 4, 7, 11, 14, 15 | 2, 5, 6, 9, 13 | 21.7 |
| 8.4 | 0, 3, 4, 7, 11, 14 | 2, 5, 9, 12, 13 | 21.6 |
| 8.5 | 0, 3, 4, 7, 10, 11, 14 | 2, 5, 9, 12 | 21.5 |
| 8.6 | 0, 3, 4, 7, 10, 11, 14 | 2, 5, 9, 12 | 21.4 |
| 8.7 | 0, 3, 4, 7, 10, 14 | 2, 5, 9, 12, 15 | 21.3 |
| 8.8 | 0, 3, 7, 10, 14 | 2, 5, 8, 9, 12, 15 | 21.2 |
| 8.9 | 0, 3, 7, 10, 13, 14 | 2, 5, 8, 12, 15 | 21.1 |
| 9.0 | 0, 3, 7, 10, 13 | 2, 5, 8, 12, 15 | 21.0 |
| 9.1 | 0, 3, 7, 10, 13 | 2, 5, 8, 11, 12, 15 | 20.9 |
| 9.2 | 0, 3, 6, 7, 10, 13 | 2, 5, 8, 11, 15 | 20.8 |
| 9.3 | 0, 3, 6, 10, 13 | 2, 5, 8, 11, 14, 15 | 20.7 |
| 9.4 | 0, 3, 6, 10, 13 | 2, 5, 8, 11, 14 | 20.6 |
| 9.5 | 0, 3, 6, 9, 10, 13 | 2, 5, 8, 11, 14 | 20.5 |
| 9.6 | 0, 3, 6, 9, 12, 13 | 2, 5, 8, 11, 14 | 20.4 |
| 9.7 | 0, 3, 6, 9, 12, 15 | 2, 5, 8, 11, 14 | 20.3 |
| 9.8 | 0, 3, 6, 9, 12, 15 | 2, 5, 8, 11, 14 | 20.2 |
| 9.9 | 0, 3, 6, 9, 12, 15 | 2, 5, 8, 11, 14 | 20.1 |
| 10.0 | 0, 3, 6, 9, 12, 15 | 1, 2, 4, 5, 7, 8, 10, 11, 13, 14 | 20.0 |
| 10.1 | 0, 3, 6, 9, 12, 15 | 1, 4, 7, 10, 13 | 19.9 |
| 10.2 | 0, 3, 6, 9, 12, 15 | 1, 4, 7, 10, 13 | 19.8 |
| 10.3 | 0, 3, 6, 9, 12, 15 | 1, 4, 7, 10, 13 | 19.7 |
| 10.4 | 0, 3, 6, 9, 12, 14 | 1, 4, 7, 10, 13 | 19.6 |
| 10.5 | 0, 3, 6, 9, 11, 14 | 1, 4, 7, 10, 13 | 19.5 |
| 10.6 | 0, 3, 6, 11, 14 | 1, 4, 7, 10, 13 | 19.4 |
| 10.7 | 0, 3, 6, 8, 11, 14 | 1, 4, 7, 10, 13, 15 | 19.3 |
| 10.8 | 0, 3, 6, 8, 11, 14 | 1, 4, 7, 10, 15 | 19.2 |
| 10.9 | 0, 3, 8, 11, 14 | 1, 4, 7, 10, 12, 15 | 19.1 |
| 11.0 | 0, 3, 5, 8, 11, 14 | 1, 4, 7, 10, 12, 15 | 19.0 |
| 11.1 | 0, 3, 5, 8, 11 | 1, 4, 7, 12, 15 | 18.9 |
| 11.2 | 0, 3, 5, 8, 11, 13 | 1, 4, 7, 9, 12, 15 | 18.8 |
| 11.3 | 0, 3, 5, 8, 11, 13 | 1, 4, 7, 9, 12, 15 | 18.7 |
| 11.4 | 0, 3, 5, 8, 13 | 1, 4, 7, 9, 12 | 18.6 |
| 11.5 | 0, 3, 5, 8, 10, 13 | 1, 4, 9, 12, 14 | 18.5 |
| 11.6 | 0, 3, 5, 8, 10, 13 | 1, 4, 9, 12, 14 | 18.4 |
| 11.7 | 0, 5, 8, 10, 13, 15 | 1, 4, 6, 9, 14 | 18.3 |
| 11.8 | 0, 5, 8, 10, 13, 15 | 1, 4, 6, 9, 14 | 18.2 |
| 11.9 | 0, 5, 10, 13, 15 | 1, 4, 6, 9, 11, 14 | 18.1 |
| 12.0 | 0, 5, 10, 15 | 1, 4, 6, 9, 11, 14 | 18.0 |
| 12.1 | 0, 5, 10, 12, 15 | 1, 4, 6, 9, 11, 14 | 17.9 |
| 12.2 | 0, 5, 7, 10, 12, 15 | 1, 4, 6, 9, 11 | 17.8 |
| 12.3 | 0, 5, 7, 10, 12, 15 | 1, 4, 6, 11 | 17.7 |
| 12.4 | 0, 5, 7, 10, 12 | 1, 4, 6, 11, 13 | 17.6 |
| 12.5 | 0, 2, 5, 7, 10, 12, 14 | 1, 4, 6, 8, 11, 13 | 17.5 |
| 12.6 | 0, 2, 5, 7, 12, 14 | 1, 6, 8, 11, 13 | 17.4 |
| 12.7 | 0, 2, 5, 7, 12, 14 | 1, 6, 8, 11, 13, 15 | 17.3 |
| 12.8 | 0, 2, 5, 7, 9, 12, 14 | 1, 6, 8, 13, 15 | 17.2 |
| 12.9 | 0, 2, 5, 7, 9, 12, 14 | 1, 6, 8, 13, 15 | 17.1 |
| 13.0 | 0, 2, 5, 7, 9, 14 | 1, 6, 8, 10, 13, 15 | 17.0 |
| 13.1 | 0, 2, 7, 9, 14 | 1, 6, 8, 10, 15 | 16.9 |
| 13.2 | 0, 2, 7, 9, 11, 14 | 1, 6, 8, 10, 15 | 16.8 |
| 13.3 | 0, 2, 7, 9, 11 | 1, 6, 8, 10, 15 | 16.7 |
| 13.4 | 0, 2, 7, 9, 11 | 1, 3, 8, 10, 12 | 16.6 |
| 13.5 | 0, 2, 7, 9, 11, 13 | 1, 3, 8, 10, 12 | 16.5 |
| 13.6 | 0, 2, 9, 11, 13 | 1, 3, 8, 10, 12, 14 | 16.4 |
| 13.7 | 0, 2, 9, 11, 13, 15 | 1, 3, 8, 10, 12, 14 | 16.3 |
| 13.8 | 0, 2, 4, 9, 11, 13, 15 | 1, 3, 10, 12, 14 | 16.2 |
| 13.9 | 0, 2, 4, 11, 13, 15 | 1, 3, 10, 12, 14 | 16.1 |
| 14.0 | 0, 2, 4, 11, 13, 15 | 1, 3, 5, 10, 12, 14 | 16.0 |
| 14.1 | 0, 2, 4, 13, 15 | 1, 3, 5, 12, 14 | 15.9 |
| 14.2 | 0, 2, 4, 6, 13, 15 | 1, 3, 5, 14 | 15.8 |
| 14.3 | 0, 2, 4, 6, 15 | 1, 3, 5, 7 | 15.7 |
| 14.4 | 0, 2, 4, 6, 8 | 1, 3, 5, 7 | 15.6 |
| 14.5 | 0, 2, 4, 6, 8, 10 | 1, 3, 5, 7, 9 | 15.5 |
| 14.6 | 0, 2, 4, 6, 8, 10, 12 | 1, 3, 5, 7, 9, 11 | 15.4 |
| 14.7 | 0, 2, 4, 6, 8, 10, 12, 14 | 1, 3, 5, 7, 9, 11, 13, 15 | 15.3 |
| 14.8 | 0, 2, 4, 6, 8, 10, 12, 14 | 1, 3, 5, 7, 9, 11, 13, 15 | 15.2 |
| 14.9 | 0, 2, 4, 6, 8, 10, 12, 14 | 1, 3, 5, 7, 9, 11, 13, 15 | 15.1 |
| 15.0 | 0, 2, 4, 6, 8, 10, 12, 14 | 1, 3, 5, 7, 9, 11, 13, 15 | 15.0 |

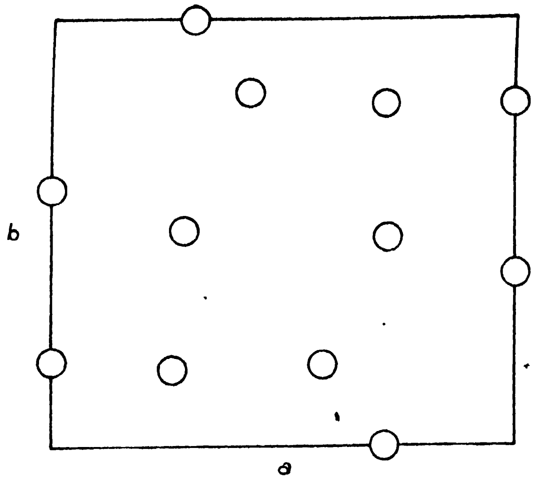


FIG. 1. Peak locations in the 001 projection of diopside determined by the approximate summation, Case I (equal amplitudes).

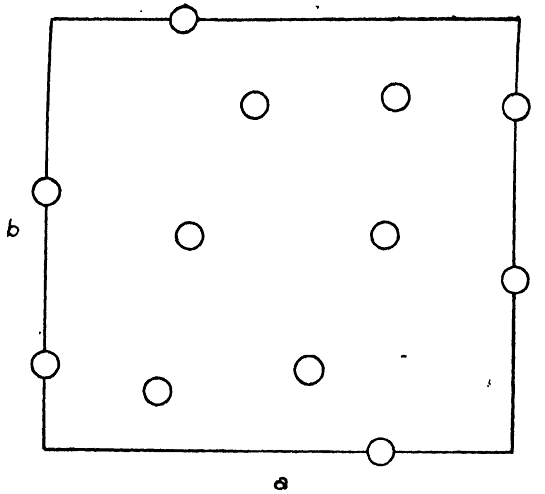


FIG. 2. Peak locations in the 001 projection of diopside determined by the approximate summation, Case II (estimated amplitudes).

Since the quarter-cell is in itself centrosymmetrical, the amount the major peaks depart from centrosymmetry will be a test of the value of the approximation. Three electron density maps have been computed. In the first, only the signs of the amplitudes have been used, all magnitudes being considered equal. In the second, estimated magnitudes have been introduced; and in the third, the actual measured values were used.

In Table I are listed the one hundred and fifty values of the constant, $C$, of Eq. (3) and the corresponding values of $x$ for which maxima and minima, as defined previously, occur. To compute the summation it is only necessary to list for each radial line the amplitudes of the reflections which are maximum and minimum at each point. As a typical example, the maxima and minima of the plane 570 along the radial line whose slope, $c$, is .4, will be determined. The constant, $C$, is equal to 5 plus .4(7), or 7.8. From Table I, opposite 7.8 it is found that maxima and minima occur at the following points:

Maxima—0, 4, 8, 11, 12, 15
Minima—2, 6, 9, 10, 13, 14

If the amplitude of the reflection, 570, is plus forty, the number forty is recorded in the columns of the worksheet corresponding to the coordinates, $x$, for which the contribution is a maximum. Similarly, minus forty is recorded in the columns corresponding to the minima. When the amplitude is negative, one must, of course, make appropriate sign changes.

## Case I—Equal Amplitudes

This example, which represents an exploration of the possibility of determining atomic locations from knowledge of the array of reciprocal lattice points, but not their relative weights, is of little

TABLE II. $F$ values of the reflections of the type $hk0$.*

| $hk0$ | $F$ | $hk0$ | $F$ |
|---|---|---|---|
| 200 | 0 | 040 | 0 |
| 400 | 15 | 240 | 16 |
| 600 | 100 | 440 | −50 |
| 800 | 76 | 640 | 29 |
| 110 | 0 | 150 | −84 |
| 310 | 76 | 350 | −55 |
| 510 | −63 | 550 | 0 |
| 710 | 41 | 750 | −96 |
| 020 | 19 | 060 | −94 |
| 220 | −51 | 260 | 31 |
| 420 | −23 | 460 | 0 |
| 620 | −19 | 170 | 0 |
| 820 | −17 | 370 | 0 |
| 130 | 13 | 570 | 40 |
| 330 | 50 | 080 | 0 |
| 530 | 37 | 280 | −32 |

* After W. L. Bragg, Zeits. f. Krist. **70**, 475 (1929).

Note: Because of symmetry, reflections of the types $\bar{h}k0$, $h\bar{k}0$, and $\bar{h}\bar{k}0$ have the same $F$ values as those of the type $hk0$.

practical importance, since one can always estimate the relative weights from a photographic film. It is included here because the results are rather striking.

The data used in this and succeeding cases were taken from Bragg[2] and are listed in Table II. The peaks of the "electron density" map drawn from the computations are represented in Fig. 1 by circles. (The actual map is, of course, distorted. In the interests of clarity, only the peak positions are shown.) Figure 1 should be compared with Fig. 4, which shows the peak locations taken from the conventional summation of Bragg.[2] It is important to note that the major peaks, representing heavy atoms or two or more atoms superimposed, are correctly positioned; small peaks are slightly displaced.

### Case II—Estimated Amplitudes

In this example, the amplitudes listed in Table II have been assigned weights of from one to five. Those from one to twenty-five, for instance, have been given the weight of one; those from twenty-six to fifty, two, etc. This was done in an attempt to simulate amplitudes that might be calculated from intensities estimated from an x-ray film. Figure 2 shows the locations of the peaks in the summation. Improvement is noted in the positions of the smaller peaks.

### Case III—Measured Amplitudes

In Fig. 3 are shown the peaks, with their relative heights indicated by contours, resulting from the approximate summation in which measured amplitudes were used. There is no appreciable change in the positions found in the previous case.

## DISCUSSION

The parameters of the peaks shown in Figs. 1 through 4 are listed in Table III. In all cases, the values obtained from the approximate summations are within .02 of those from the conventional summation. (The values for the conventional summation may or may not be in agreement with the actual parameters of diopside. Those listed were measured from the electron density map in order to present a fair comparison with those derived in the same manner from the approximate summations.) On the basis of this one test, it would appear that the electron density peaks can, indeed, be located with reasonable accuracy by considering only the maximum and minimum contributions at each point in the cell. One would, however, hesitate to recommend the approximate method until independent checks have been made. The choice of diopside may have been a fortuitous one.

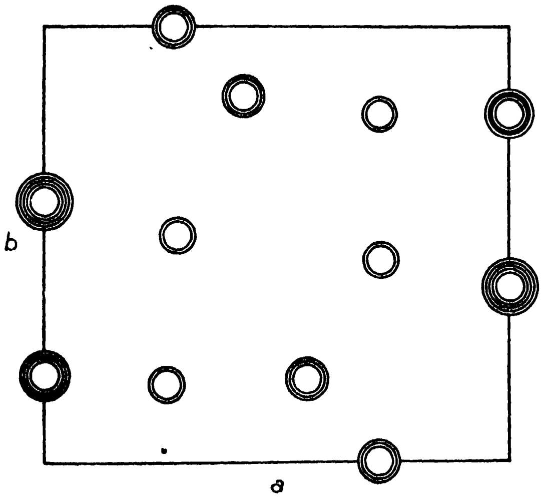


FIG. 3. Peak locations in the 001 projection of diopside determined by the approximate summation, Case III (measured amplitudes).

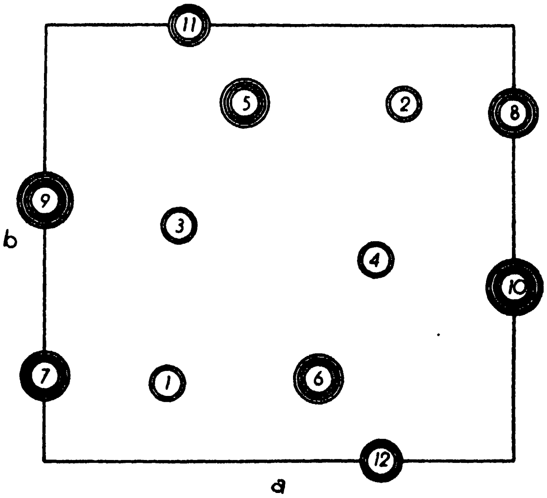


FIG. 4. Peak locations in the 001 projection of diopside determined by the conventional summation (after Bragg).

TABLE III. Parameters measured from Figs. 1 through 4.

| Peak number | | Case I (no amplitudes) | Case II (estimated amplitudes) | Case III (measured amplitudes) | Complete summation (after Bragg) |
|---|---|---|---|---|---|
| 1 | *a* | .12 | .13 | .13 | .13 |
| | *b* | .07 | .09 | .09 | .09 |
| 2 | *a* | .37 | .36 | .36 | .38 |
| | *b* | .41 | .40 | .40 | .41 |
| 3 | *a* | .15 | .14 | .14 | .14 |
| | *b* | .25 | .25 | .26 | .27 |
| 4 | *a* | .36 | .36 | .36 | .35 |
| | *b* | .25 | .24 | .23 | .23 |
| 5 | *a* | .22 | .21 | .21 | .21 |
| | *b* | .40 | .41 | .42 | .41 |
| 6 | *a* | .28 | .29 | .28 | .29 |
| | *b* | .09 | .09 | .09 | .09 |
| 7 | *a* | 0 | 0 | 0 | 0 |
| | *b* | .10 | .10 | .10 | .10 |
| 8 | *a* | .50 | .50 | .50 | .50 |
| | *b* | .40 | .40 | .40 | .40 |
| 9 | *a* | 0 | 0 | 0 | 0 |
| | *b* | .30 | .30 | .30 | .30 |
| 10 | *a* | .50 | .50 | .50 | .50 |
| | *b* | .20 | .20 | .20 | .20 |
| 11 | *a* | .14 | .15 | .14 | .15 |
| | *b* | .50 | .50 | .50 | .50 |
| 12 | *a* | .35 | .35 | .35 | .35 |
| | *b* | 0 | 0 | 0 | 0 |

One of the potential uses of the approximate method is in the checking of an assumed structure. If a logical structure can be deduced from crystal chemical knowledge, the phases of the strong reflections can be computed. An approximate summation, using only those reflections, might suffice to demonstrate whether or not the assumed structure is consistent.

## COMPUTATION BY MEANS OF STRIPS

Although the labor of computation is reduced in the approximate method to the addition of columns of numbers, it is still necessary to look up data in a table and to record them. This effort can be eliminated by preparing a series of strips, one for each value of the constant, *C*. Each strip has a space for each value of the coordinate, *x*, and those spaces where the function is a maximum are colored blue. Minima are indicated by red. The labor of summation involves only the selection of the strips, their arrangement on a suitable board, and the addition of the coefficients.

In checking an assumed structure, it is not necessary to perform the entire summation. Only those regions which should be high in density need be treated. Using the approximate method, the summation can be done in a very short time.

# Single Crystal Electron Diffraction by Micro-Crystalline Materials*

NORMAN DAVIDSON** AND JAMES HILLIER
*RCA Laboratories, Princeton, New Jersey*

(Received March 11, 1947)

It is shown that the double lens probe-type electron diffraction camera described earlier is capable of producing useful electron diffraction patterns of single microscopic crystals in the size range 200–2000A; the formation of an opaque carbonaceous deposit, which occurs when an electron probe is focused on a specimen, is prevented in the diffraction camera by previously exposing the specimen chamber to an electron spray. A number of examples of two-dimensional grating patterns and Kikuchi-line patterns obtained from such crystals are shown. From either type of pattern it is possible to obtain accurate values for some of the lattice spacings, while from the Kikuchi-line patterns it is possible to deduce accurately the orientation of the crystals relative to the electron beam. The results are shown to agree reasonably well with the predictions of the dynamical theory as presented by MacGillavry. It is concluded that a two-dimensional grating pattern is to be expected from a crystal oriented with an important lattice vector approximately parallel to the electron beam independent of slight disorders or warping of the crystal. Kikuchi-line patterns are to be expected for most other crystals which are not excessively thick or disordered. The theory also shows that both types of patterns may be obtained simultaneously.

## I. INTRODUCTION

THE study of the diffraction of fast electrons by single crystals has been pursued in several different ways up to date. Because of the limited penetration of the electron beam, work with macroscopic crystals which are readily manipulated and oriented is restricted either to reflection diffraction by a surface or, for transmission electron diffraction, to the relatively rare case of materials, such as mica, which can be obtained as thin sheets.

There would be many useful applications of an effective technique for obtaining single crystal electron diffraction patterns and thus determining the lattice constants of crystals of dimensions less than a micron—that is, crystals of dimensions appropriate to electron diffraction and electron microscopy. Some of these applications are:

(a) Identification of a material which has a complex powder pattern or which is present as a minor constituent of a mixture of crystals.

(b) The determination of crystal structure when single crystals of sufficient size for x-ray diffraction are not available.

(c) Further study of the fundamental phenomena of electron diffraction by crystals.

Various methods of observing the electron diffraction patterns from single micro-crystals have been reported in the literature. The observations of crystalline reflections in electron micrographs[1,2] have been used to confirm the dynamical theory of electron diffraction, to study the warping of crystals, and in general to provide information about strong reflections by properly oriented crystals. The method is limited to strong reflections.

Boersch has obtained electron diffraction patterns in which the collimated electron beam was restricted by an aperture to have a cross section of 0.005 mm at the specimen.[3] The spot patterns obtained for the various films examined were such that only one or a few crystals were responsible for the diffraction, but no correlation was possible between the morphology of the irradiated portion of the specimen and its diffraction pattern.

The same author has suggested a scheme in which a fine aperture moves in the image plane of an electron microscope.[4] When the aperture

---

* Presented before the American Society for X-Ray and Electron Diffraction, June 10, 1946, at Silver Bay, New York.

** Present Address: Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena, California.

[1] R. D. Heidenreich and L. Sturkey, J. App. Phys. 16, 97 (1945).

[2] J. Hillier and R. F. Baker, Phys. Rev. 61, 722 (1942). References 1 and 2 refer to other papers on this topic.

[3] H. Boersch, Zeits. f. Physik 116, 469 (1940).

[4] H. Boersch, Ann. d. Physik 27, 78 (1936).

is placed on the image of a particular crystal, the angular distribution of the transmitted rays provides information about the diffraction pattern of the particular crystal. Boersch used this technique to obtain diffraction patterns of 0.08-mm portions of gold foil. This method would be difficult to develop into a practical method for true micro-crystals because of the effects of spherical aberration on the diffracted rays, and of the decrease, caused by magnification, in angular aperture of an imaging pencil.

Ardenne has suggested several schemes for modifying the electron optical system of an electron microscope so as to obtain diffraction patterns from fields 1–10 microns in diameter.[5, 6]

The purpose of the present article is to report the phenomena observed using a mode of operation of the electron diffraction camera described by Hillier and Baker[7] wherein a beam of electrons reduced to a probe of *ca.* 200A cross section, and of maximum angular aperture *ca.* $7.5\times10^{-3}$ radian is diffracted by single micro-crystals. It is believed that this method offers a practical approach to the problem of obtaining electron diffraction patterns from single micro-crystals, and of correlating the morphology and orientation of a particular crystal with its diffraction pattern.

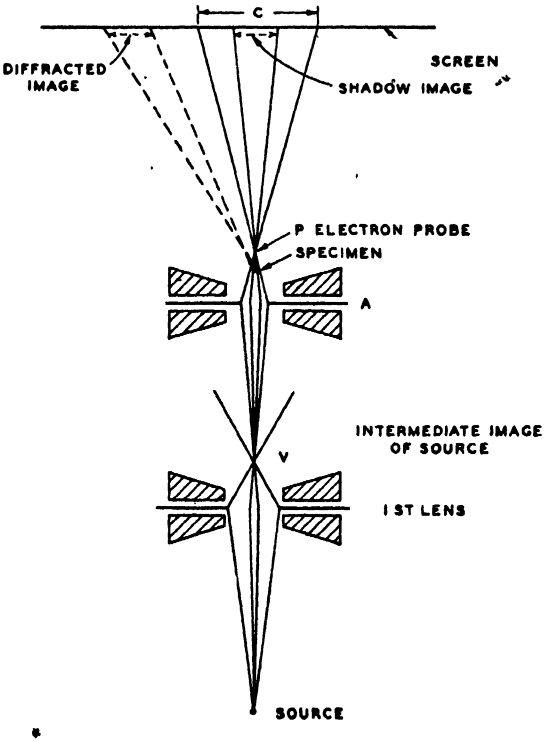


FIG. 1. Schematic diagram of lens system and ray paths in the electron diffraction camera as used for shadow microscopy and single crystal diffraction.

## II. EXPERIMENTAL TECHNIQUES

The double lens system of the electron diffraction camera can be used in the two following ways (Fig. 1):

(a) Using the first lens as a strong lens to form a reduced image, *V*, of the electron source, and the second lens to re-focus the image on the photographic plate or fluorescent screen, a focused electron diffraction pattern is obtained of an area of specimen determined by the size of the limiting aperture, *A*, of the second lens.

(b) When the focal length of the second lens is reduced, the image *V* of the source is re-focused to an electron probe *P*, somewhere between *A* and the image screen. A bright field shadow image of the specimen is formed on the image screen within the central cone of illumination, *C*. The rays diffracted out of *C* by crystalline material in the irradiated field are the subject matter of this article.

From the point of view of obtaining a conventional powder diffraction pattern, one is producing, in this mode of operation, a "defocused" electron diffraction pattern. The degree of defocusing may be expressed by the numerical value of the angular aperture of the convergent beam of electrons passing through crystalline material in the field of irradiation. It will be seen, however, that under some circumstances it is possible to obtain precise values of interplanar spacings and lattice constants from defocused patterns.

For the sake of clarity, some of the well-known facts about the operation of a shadow microscope[8, 9] should be mentioned before describing

---

[5] M. von Ardenne, Zeits. f. Physik **117**, 515 (1941).

[6] M. von Ardenne, E. Schiebold and F. Gunther, Zeits. f. Physik **119**, 352 (1942).

[7] J. Hillier and R. F. Baker, J. App. Phys. **17**, 12 (1946).

[8] These factors are discussed in some detail and other references given in reference 7.

[9] V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, and A. W. Vance, *Electron Optics and the Electron Microscope* (John Wiley and Sons, New York, 1945), Chapters III and XIX.

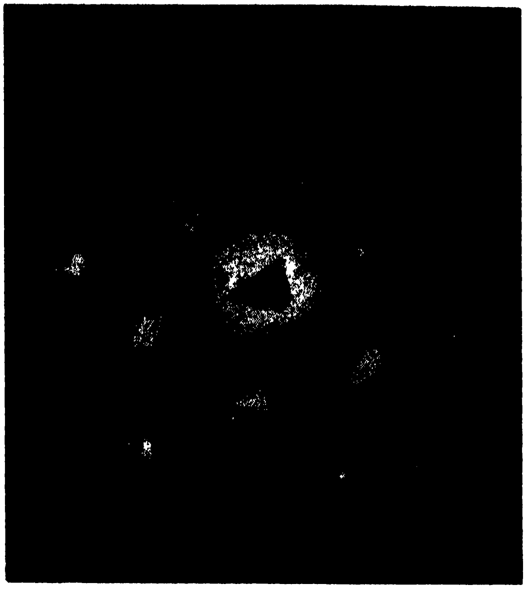


FIG. 2. Two-dimensional grating pattern from graphite, *kho*, reproduced from Hillier and Baker, reference 7, Fig. 15 (c). The bright field shadow image of the crystal has been printed on from a negative given a shorter exposure. Note the variation of reflecting power of different areas of the crystal for different reflections.

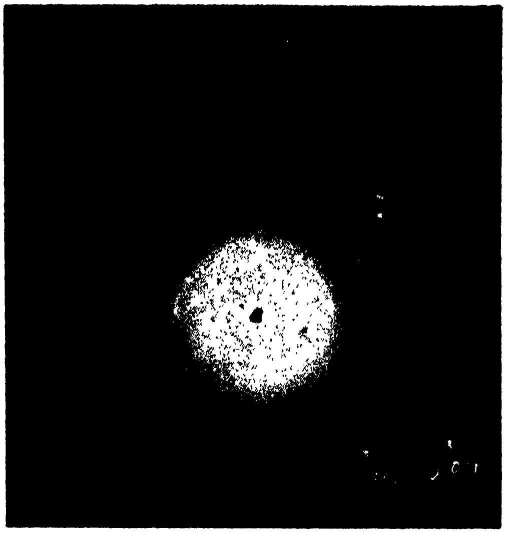


FIG. 3. Two-dimensional grating pattern, BeO (hexagonal), *hol*, $d_{100}=2.33$A, $d_{001}=4.38\pm0.02$A. (ASTM, 2.34A, 4.38A, respectively.) The bright field shadow image of the crystal has been printed on from a negative given shorter exposure. The $10\bar{4}$ reflection is relatively much brighter than for the x-ray case; the $\bar{3}01$ and $\bar{3}03$ are evident, although the calculated structure factor for these reflections is zero.

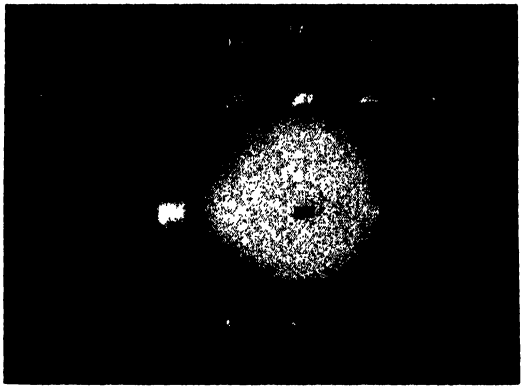
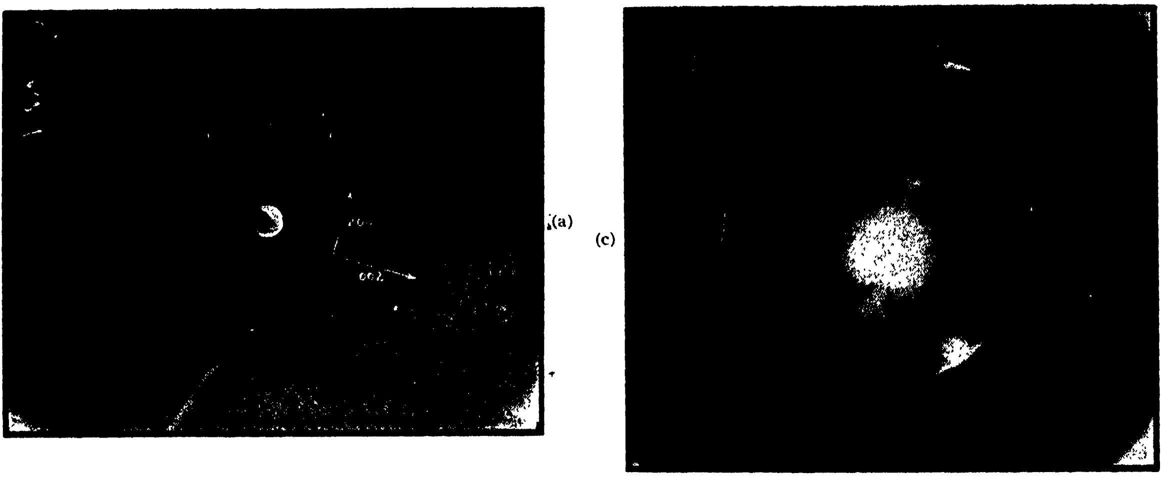


FIG. 4. BeO, *hhl*, $d_{110}=1.33_5\pm0.01$A, $d_{002}=2.20_5\pm0.01$A. (ASTM, 1.350A, 2.19A.) Note the intensity of the *hh*4 and $hh\bar{4}$ reflections. The central image of the crystal has been printed on from a shorter exposure.

the diffraction results. The magnification and the resolving power of a shadow microscope are dependent on the probe position with respect to specimen and electron-recording screen and the probe size. To obtain a small probe, the focal length of the lens should be as short as possible. To obtain high magnification the probe should be reasonably close to the specimen. In practice a probe of minimum cross section of *ca.* 100–200A is obtainable with lenses of 2–3 mm focal length. While no precise tests of the resolving power of the present instrument have been made, resolutions less than 1000A and probably somewhat more than 200A have been obtained with magnifications of from $4\times10^3$ to $10^4$. With the pole piece configuration employed the focal point of the second lens for a probe of *ca.* 200A cross section was beyond the lens opening; it was, therefore, easy to mount the specimen close to the probe position. Under these conditions and with a 25-$\mu$ aperture for $A$, the maximum angular aperture of irradiation of the specimen was *ca.* $7.5\times10^{-3}$ radian. The angular aperture decreases rapidly as the separation of the probe and specimen is increased, and was $10^{-3}$ for most of the work described below.

When the probe is focused directly on the specimen, the magnification is infinite and no specimen image is observed, but rather a projection of the limiting aperture, $A$, provided the specimen is partially transparent. Actually, be-



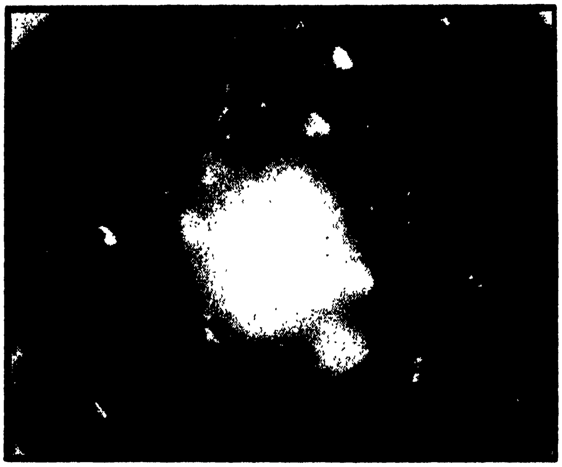


FIG. 5. Tip of $\alpha Al_2O_3 \cdot H_2O$ crystal (böhmite) (orthorhombic) $2h02l$, $d_{100} = 1.86 \pm 0.01A$, $d_{002} = 1.42 \pm 0.01A$. (ASTM, 1.86A, 1.43A.) (a) Short exposure showing image of crystal and one strong diffracted image, $\bar{2}02$. (b) Longer exposure showing full two-dimensional pattern and some $2h+1$, ?, $2l+1$ reflections. (c) Probe focused on specimen; each diffraction spot is an image of the defining aperture, some containing the striations discussed in the text. The Kikuchi-line pattern shows that the strongly diffracting $\bar{2}02$ plane is oriented at its Bragg angle with respect to the incident beam.

cause of asymmetries in the lens,[10] the image will first go to infinity in one direction, and then in another perpendicular to the first as the focal power is varied. Images observed at near infinite magnification were somewhat distorted because of this asymmetry.

The correlation of diffraction pattern with morphology and orientation of the crystal can be made on the basis of the electron shadow micrograph. Alternately, the shadow micrograph may be used for control purposes. With a little practice, the same crystal can be located in the field of view of a conventional magnetic electron microscope, which provides a higher resolution image of the crystal. One difficulty with the electron shadow microscope as used here is that the magnification, which varies with the exciting current of the magnetic electron lens, the position of the pole-piece material on its hysteresis loop, and the specimen-lens distance, was unknown for any particular photograph. It is estimated that most of the crystals shown here for which conventional micrographs are not shown were *ca.* 2000A–5000A in dimensions. The $\alpha Al_2O_3H_2O$ crystals are usually 5–10 microns long.

### III. EXPERIMENTAL RESULTS

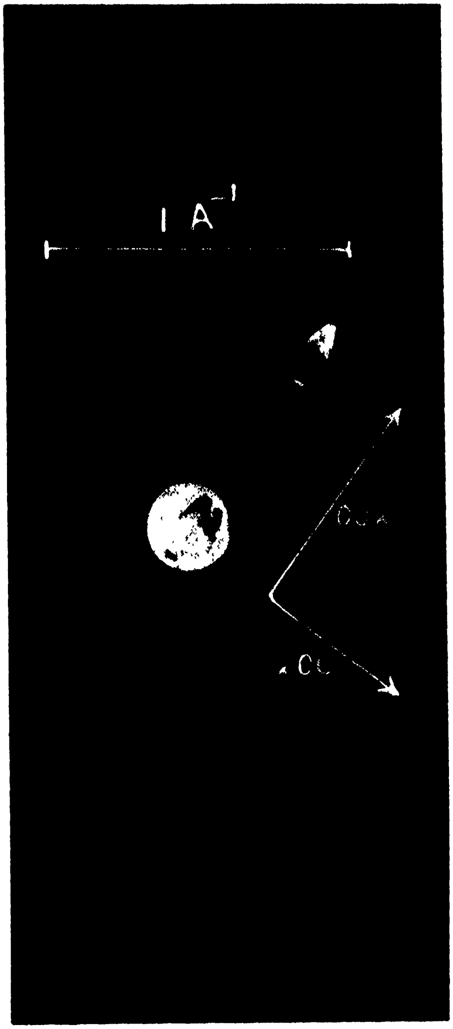
The diffraction results are very similar to those observed by Kossel and Möllenstedt for the diffraction of a convergent beam of electrons by thin sheets of mica.[11] The present work is an extension of their results to single crystal analyses of *arbitrary* micro-crystals.

Two principal types of diffraction patterns were observed under the experimental conditions described above: (a) two-dimensional grating patterns and (b) Kikuchi-line patterns. Figures 2–7 are illustrations of two-dimensional grating

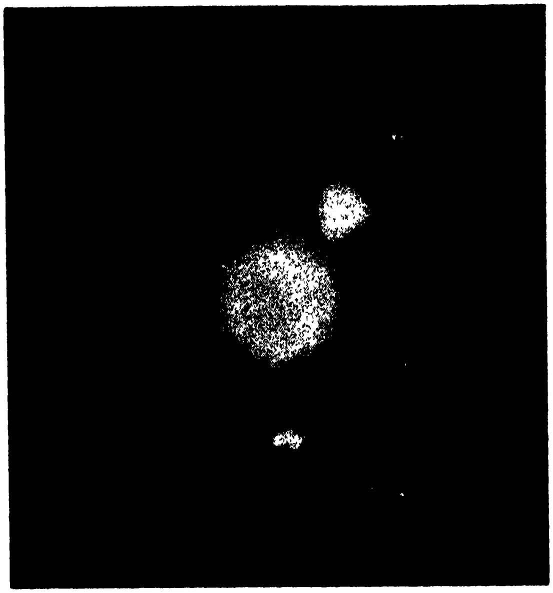
---

[10] J. Hillier, J. App. Phys. 17, 307 (1946).

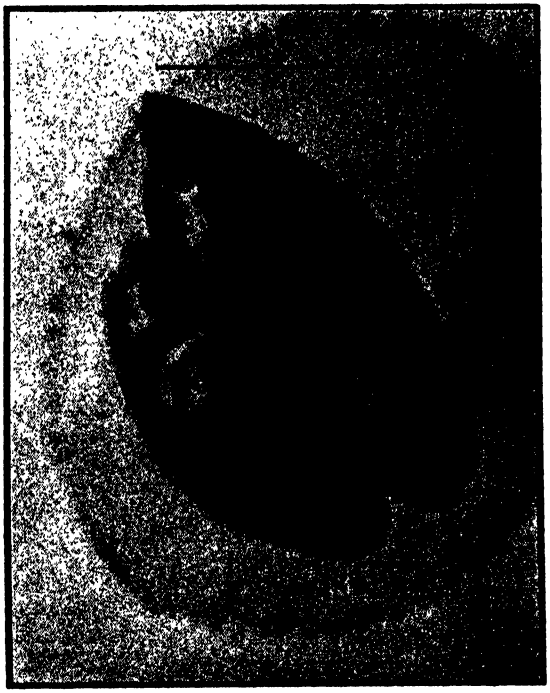
[11] W. Kossel and G. Möllenstedt, Ann. d. Physik 36, 113 (1939).



(a)



(b)



(c)

FIG. 6. Diffraction pattern from a pair of $Al_2O_3 \cdot H_2O$ crystals. In the short exposure (a), there is one strong diffracted image (002) from one of the crystals. One portion of this crystal diffracts much more strongly than the other. It is evident from the position of the Kikuchi line in the longer exposure (b) that the strongly diffracting region of the crystal makes the Bragg angle with the incident beam. A two-dimensional grating pattern is developing in (b) and became much more evident in longer exposures not shown here. Because the two crystals are not exactly parallel, the diffracted images are separated. Note the variation in reflecting power of neighboring parts of the crystal for a particular reflection, especially in the 002 and $20\bar{2}$ reflections. This was noted in the other images on longer exposure. An electron micrograph of the same pair of crystals is shown, for comparison, in (c). Because of the large field of view required to include both crystals, the shadow micrograph in (a) was not obtained with the highest usable magnification and the smallest probe size, and is not intended to illustrate the ultimate in performance of the shadow microscope.

patterns. In these cases, the probe was focused between the specimen and the electron-recording screen, so that a bright field shadow image of the crystal is formed in the central cone of illumination.

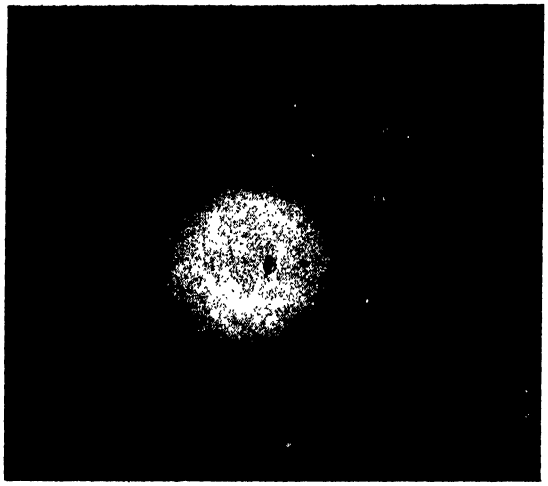


FIG. 7. ZnO (hexagonal), *hh*2, $d_{110} = 1.62 \pm 0.01$A, $d_{002} = 2.61 \pm 0.02$A. (ASTM, 1.61A, 2.61A.)

The field of view was adjusted so that only one crystal was present. Through any one point of the crystal passes a pencil of rays of very small angular aperture (*ca.* $10^{-4}$ radian, depending on the probe size and the specimen to probe separation). If only the rays passing through one portion of the crystal were diffracted, there would be an uncertainty in the diffraction angle proportional to the dimensions of the shadow image. Actually it is frequently observed in the two-dimensional grating patterns that each diffracted spot is a dark field shadow image of the diffracting crystal. Furthermore, the distances between corresponding points of a diffracted image and the central image correspond precisely to the angles of scattering deduced from the interplanar spacings of the crystal; the azimuthal directions of scattering are those of the normals to the corresponding planes of the crystal. The two-dimensional grating patterns may be considered as being formed by linear combinations (with integer coefficients) of two vectors, $\mathbf{b}_{hkl}$ and $\mathbf{b}_{h'k'l'}$, of the reciprocal lattice of the crystal. The total deviation of any ray that is diffracted must be twice the Bragg angle for a particular reflection.

When such two-dimensional grating (cross-grating) patterns are obtained, there is an apparent breakdown of the Bragg law for the diffraction of electrons by the crystal. Although the total deviation of a ray is invariably twice the Bragg angle for a particular reflection, the beam need not be incident on a crystal at precisely the Bragg angle with respect to the reflecting plane. It is sufficient that the ray be approximately parallel to the reflecting plane.

It is evident that in order for a cross-grating pattern to be possible, two of the shorter vectors of the reciprocal lattice must be approximately perpendicular to the electron beam; an equivalent statement is that an important lattice vector must be approximately parallel to the electron beam.

The method of specimen dispersion generally used[12]—spreading a drop of an amyl acetate solution of collodion containing a fraction of a milligram of the powdered specimen between two microscope slides by sliding them against each other—favors orientation of the crystals with well developed faces parallel to the collodion film. For the crystals used, this condition is equivalent to having an important lattice vector approximately parallel to the electron beam.

It should be emphasized that the intensity diffracted onto most of the spots observed in the cross-grating patterns is quite low—from 0.1 percent to 3 percent of the incident intensity. Sometimes one or two of the reflections are very strong containing practically all of the scattered intensity. The crystal is presumably oriented so that the beam makes an angle close to the Bragg angle with the plane responsible for these reflections. Figures 5 and 6 provide examples of this phenomenon.

As will be discussed in more detail later on, calculations based on the dynamical theory of electron diffraction indicate that diffracted intensities of the order of magnitude of from 0.1 percent to 3 percent are to be expected for each of a set of spots of a two-dimensional grating pattern when an electron beam is incident approximately parallel to an important lattice vector of a perfect crystal. In addition, any disorder or warping of the crystal will contribute to the development of a two-dimensional grating

---

[12] This technique was used in all the examples cited in this article except for the case of $\alpha Al_2O_3 \cdot H_2O$. An aqueous suspension of this material dries down on a collodion film with most of the crystals lying flat on the film.

pattern, a factor which is probably important for most of the patterns exhibited here.

Qualitatively, it is observed that the relative intensities of the various diffraction spots of a two-dimensional grating pattern are not the same as the x-ray diffraction intensities for the given material. In particular, reflections which are very weak for the x-ray case may have a much greater relative intensity for electron diffraction. Of course, these anomalous intensities have been observed frequently before in electron diffraction patterns.[13] A reflection which would be weak or vanishing by the structure-factor criterion, but which can be formed by the vector sum of two strong reflections will have appreciable intensity because it will be formed by the process of multiple scattering. This is possible for electron diffraction because a beam need not be incident at the Bragg angle with a plane to be strongly diffracted by that plane.

The most striking examples of this phenomenon of anomalous intensities among the samples studied occurred for BeO. The *h*04 reflections of Figs. 3 and 4 are as strong as the other reflections, although the *hk*4 reflections of BeO are extremely weak in the case of x-ray diffraction.[14] The 301 and 303 reflections are weak but evident (Fig. 3) although the structure factor for these reflections is zero. The 301 reflection, for example, may be considered as resulting from the combination of 200 and 101, or 300 and 001, etc., all of which are allowed and strong. However, in Fig. 4 there are only *hh* 2*l* reflections, and no *hh* 2*l*+1. In this case there is no way of combining two allowed reflections to obtain a forbidden reflection.

It should be recalled at this point that Bethe in his fundamental paper on the dynamical theory of electron diffraction[15] suggested the possibility of these "anomalous" reflections, resulting from the combination of two strong reflections.[16]

In the x-ray diffraction patterns from $\alpha Al_2O_3 \cdot H_2O$, the symmetry of the orthorhombic crystal prohibits all but the 2*h* 0 2*l* reflections of the *h*0 *l* series.[17] However, in longer exposures images appear in the middle of the rectangles that form the orthorhombic pattern. The lattice constants of $\alpha Al_2O_3 \cdot H_2O$ are $a = 3.72$A, $b = 12.5$A, $c = 2.83$A. Because of the large value of *b*, the *h*1*l* reflections occur through nearly the same angle, and at approximately the same angle of incidence as the *hol* reflections. Therefore, the indices of the "anomalous" spots are probably 2*h*+1, 1, 2*l*+1.

The second general class of phenomena observed are the production of Kikuchi-line patterns. In general, three requirements must be fulfilled in order for the Kikuchi-line phenomena to occur:

(a) The crystal must be sufficiently thick so that the primary process of incoherent scattering—to produce a cone of rays of much larger angular aperture than the incident cone of illumination—take place with sufficient probability to produce a general blackening of the entire photographic plate. In order that all the background scattering should contribute to the patterns, the electron probe is focused on the crystal in these experiments.

(b) The Bragg law must be satisfied for the diffraction of electrons by the crystal; that is, electrons incident upon a reflecting plane at the Bragg angle must be diffracted with much greater intensity than electrons incident at another angle.

(c) Some planes must be oriented approximately parallel to the axis of the electron beam. This requirement will usually be fulfilled because of the great number of reflecting planes in a crystal.

The distance from a "depleted" line to the corresponding "enhanced" line of a Kikuchi pattern corresponds to an angle of scattering of twice the Bragg angle. The positions of the various lines of a Kikuchi-diffraction pattern allow one to infer not only some of the interplanar spacings of the crystal, but also the precise orientation of the crystal with respect to the electron beam. Where a "depleted" line runs through the central area of irradiation, some of the rays of the incident radiation must make an angle equal to the Bragg angle with this particular plane, and the entire cone of illumination makes an angle quite close to the Bragg angle. There will then be a weak diffracted image of the defining aperture containing the strongly dif-

---

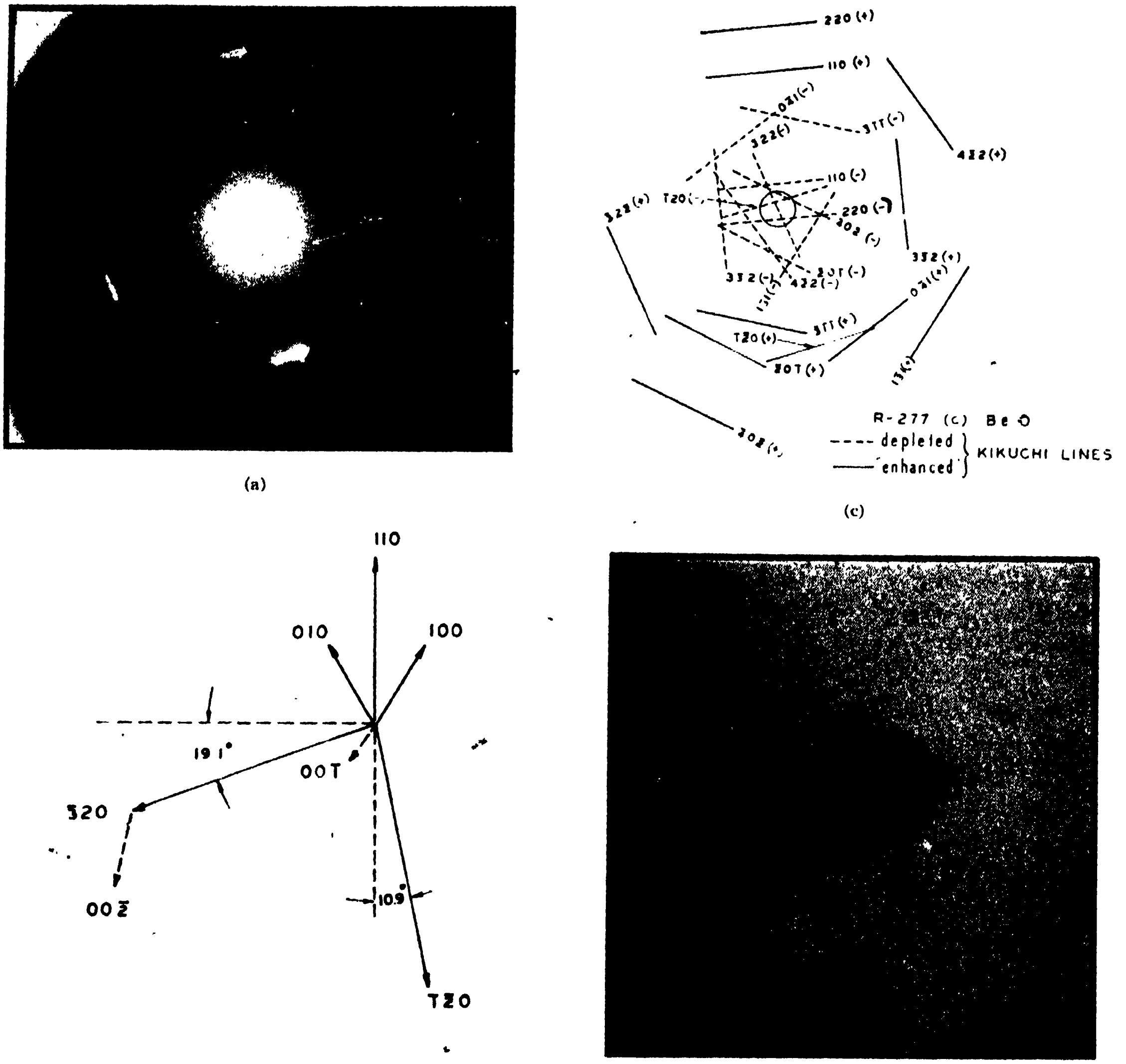
[13] e.g., L. H. Germer, Phys. Rev. **61**, 309 (1942).

[14] W. H. Zachariasen, Zeits. f. Physik **40**, 637 (1927), lists as the relative intensities of the 002 and 004 reflections for BeO, 2.5 and 0–0.5.

[15] H. Bethe, Ann. d. Physik **87**, 55 (1928).

[16] *Ibid.*, pp. 97–99.

[17] S. Goldstaub, Bull. Soc. Franc. Mineral. **59**, 348 (1936). See also C. Gottfried, Strukturbericht, **4**, 129 (1936).



(a)

(c)

(d)

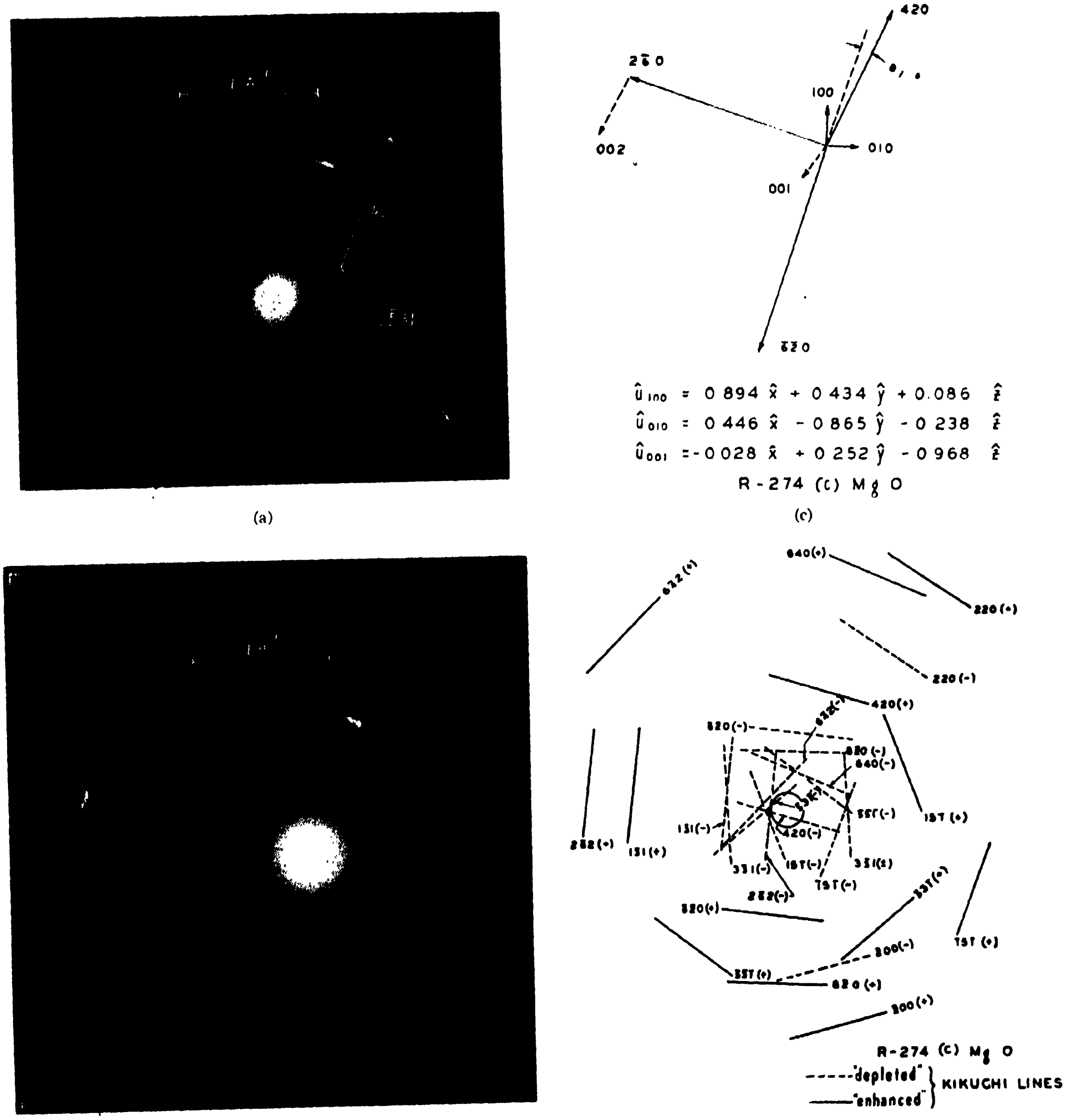
$$\bar{b}_{100}=3.89\,\hat{x}-2.67\,\hat{y}-1.00\,\hat{z}$$
$$\bar{b}_{010}=4.42\,\hat{x}+1.82\,\hat{y}+0.68\,\hat{z}$$
$$\bar{b}_{001}=\quad -0.90\,\hat{y}+2.41\,\hat{z}$$
$$11.24\,\mathrm{mm}=1\,\mathrm{A}^{-1}$$

R-277 (c) Be O

(b)

FIG. 8. (a) Kikuchi-line pattern from a BeO crystal, with diffracted images of aperture of illumination. Note that there is a 110-diffracted image and a broad 110-Kikuchi line that does not go through this image. (b) Deduced orientation of the reciprocal vectors of the BeO lattice with respect to axes in space. The electron beam is moving in the direction of the $+z$-axis. (c) Calculated Kikuchi-line pattern from the transformation equation. All the calculated lines are present in the photograph except in cases (such as $3\,\bar{1}\,\bar{1}$ or $0\,\bar{4}\,1$) where the + and − lines are almost symmetrically disposed with respect to the central unscattered beam. (d) An electron micrograph of the crystal responsible for this pattern. It seems likely that the crystal is oriented with the hexagonal axis perpendicular to the collodion film; because of proximity to the supporting mesh, the film is not perpendicular to the electron beam direction.

fracted beam caused by the rays that were incident at the Bragg angle. In the circular images, there may be a set of striations perpendicular to the direction from the central image to the diffracted image. (These parallel striations will be bent and otherwise perturbed when another Kikuchi line runs through them.) This



(a)

(c)

(b)

(d)

Fig. 9. (a) and (b) show the Kikuchi-line pattern from an MgO micro-crystal. The deduced orientation is exhibited in (c) and (d). In this case, the unit vectors perpendicular to the faces, rather than the reciprocal vectors, are used in the transformation equations. The specimen was prepared by dispersing MgO smoke in collodion; this procedure promoted the isolation of single MgO crystals.

Note the difference in the (420) striation patterns of (a) and (b), two different exposures, indicating a very slight change in orientation between exposures. The calculated and observed Kikuchi patterns are in agreement except that the reflections with odd indices are absent except for the case of ($\bar{5}$ 3 $\bar{1}$), where the plane is at its Bragg angle with respect to the incident beam.

phenomenon was observed for mica by Kossel and Möllenstedt, and interpreted by MacGillavry[18] in terms of the dynamical theory of electron diffraction. The spacings of these striations permit a determination of the thickness of the crystal (assuming that it is of uniform thickness in the direction of the electron beam), and

---

[18] C. H. MacGillavry, Physica 7, 329 (1940).

in favorable cases, of the structure factor for the particular reflection.

Figures 8, 9 and 5c are examples of the Kikuchi lines produced when the electron probe is concentrated on a micro-crystal. Calculated patterns and the inferred orientation of the crystal are also illustrated for Figs. 8 and 9. The orientations were deduced by requiring that two of the planes whose negative Kikuchi lines run through the central cone of illumination make an angle equal to their respective Bragg angles with the $z$-axis (the direction of the electron beam in Figs. 8b and 9c). The predicted patterns are calculated by finding by inspection all the linear combinations (with integer coefficients) of the unit reciprocal vectors that have small negative components along the $z$-axis. In all cases, the measured spacings and angles between Kikuchi lines were correct within experimental errors (±0.2–5 percent for the spacings, depending on the sharpness of line; ±*ca.* 1°–2° for the angles, depending on the length and sharpness of the lines).

In the early stages of the present work, when the electron probe was focused on a specimen or a membrane, a contamination "spot" formed rapidly on the irradiated area. Figure 6 (c) is a conventional electron micrograph of a pair of $Al_2O_3 \cdot H_2O$ crystals that had been exposed to an electron beam of rather large area for several minutes in the diffraction unit; the contamination spot, defining the area of irradiation, approximately 2.5 microns in diameter, is clearly visible. In practice, the area irradiated with a concentrated probe would become completely opaque to electrons after 30–60 seconds of bombardment. It was not possible, therefore, to study systematically the production of Kikuchi lines by micro-crystals until a method of preventing contamination was developed.

This difficulty in "probe" analysis was previously encountered in the electron microanalyzer.[19] The carbonaceous nature of the deposit is demonstrated in this instrument; the inelastic scattering spectra of thin films of silver or silica[20] showed no peak caused by the $K$-level of carbon when first exposed to a probe of *ca.* 2000A diameter. In later exposures the carbon $K$-line developed, and by the time the specimen was opaque only the carbon line was observed.

It has been discovered that, in the diffraction unit or in the electron microanalyzer, contamination can be avoided if the specimen chamber of the unit is exposed to an electron spray (60ma at 800v) for a few minutes before concentrating the probe on the specimen. Contamination will be entirely negligible for 10 minutes or so, and not serious for several days after this treatment. It should be mentioned that this technique is not effective in preventing contamination of electron microscope specimens. Here contamination appears to be a somewhat more complicated phenomenon.

In early stages of the work, before the method of avoiding contamination was discovered, patterns of a few spots were frequently obtained when the probe was focused behind or in front of the specimen. If the diffracted spot does not form a complete image of the crystal, and if a complete two-dimensional array is not obtained, it is difficult to interpret the pattern. One reason for this is that the particular portion of the crystal that is diffracting the beam is unknown, so that a precise measurement of the angle of scattering is not feasible. It is believed, however, that most of such crystals would give useful Kikuchi patterns which provide precise information about the inter-planar spacings and angles.

In concluding the description of the experimental work, it should be pointed out that the work to date has been confined to the study of the patterns produced by crystals of known structure. It is evident in the Kikuchi patterns of Figs. 8 and 9 that even a highly symmetrical and simple crystal can give rise to a quite complicated electron diffraction pattern. The application of the method to the identification of unknown crystals or the determination of their lattice constants, would be greatly facilitated by the development of a stage which allows precise adjustment of the orientation of the crystal with respect to the electron beam without moving it from the field of view.

## IV. THEORETICAL INTERPRETATION

The spacings of the striations in the 420 reflection of MgO in Fig. 9(a) were compared with the predictions based on the dynamical

---

[19] J. Hillier and R. F. Baker, J. App. Phys. 15, 671 (1944).

[20] Silica experiment by R. F. Baker.

TABLE I. Observed and calculated displacements between maxima of striations of Fig. 9a.

| Maxima involved | $\bar{1}$–0 | 0–1 | 1–2 | 2–3 | 3–4 | 4–5 |
|---|---|---|---|---|---|---|
| Observed displacement ($\times 10^4$) | 13.0 | $13.4_5$ | 7.3 | 7.8 | 6.9 | 5.9 |
| Calculated displacement | 13.1 | 13.1 | 7.9 | 7.4 | 7.0 | 7.0 |

theory of electron diffraction. The formula given by MacGillavry[18] for the angular variation of diffracted intensity is:

$$\frac{I(\Delta)}{I_0}=\frac{\sin^2\left\{\frac{1}{2}\frac{Dv_{hkl}}{K}\left(1+\frac{4\pi^2\Delta^2K^2}{d_{hkl}{}^2v_{hkl}{}^2}\right)^{\frac{1}{2}}\right\}}{1+\frac{4\pi^2\Delta^2K^2}{d_{hkl}{}^2v_{hkl}{}^2}},\tag{1}$$

where $\Delta$ = deviation of angle of incidence from Bragg angle, in radians

$K=2\pi/\lambda$ = wave vector of electron beam,

$D$ = thickness of crystal,

$d_{hkl}$ = inter-planar spacing,

$$v_{hkl}=\frac{8\pi^2me}{h^2}\frac{d^2e}{\pi R}\sum\nolimits_j(z_j-f_j)\exp\{2\pi i(\bar{b}_{hkl}\cdot\bar{r}_j)\}.$$

In this last relation, $R$ is the volume of the unit cell, and $\sum_j$ the electronic structure factor for the *hkl* reflection. The total deviation of the rays in this treatment is always $2\theta_B$.

The formula (1) is applicable to a crystal slab of uniform thickness in the direction of the electron beam. It is assumed that only one strong reflection occurs. There are clearly two strong reflections in Fig. 9. In Fig. 9(a), the striations are sensibly parallel and perpendicular to the direction of scattering so that this pattern was chosen for analysis. It is interesting that in the next exposure (a minute or so later) (Fig. 9(b)) the 420 pattern was highly perturbed by interaction with the $\bar{2}62$ reflection; evidently a very slight change in the orientation of the crystal took place. The next exposure showed parallel striations again. If in formula (1) we set $\Delta_l=d/D$,

$$\frac{I(\Delta)}{I_0}=\frac{\sin^2\left\{\pi\left(\frac{\Delta^2}{\Delta_l{}^2}+\frac{d_{hkl}{}^2v_{hkl}{}^2}{4\pi^2\Delta_l{}^2K^2}\right)^{\frac{1}{2}}\right\}}{1+\frac{4\pi^2\Delta^2K^2}{d_{hkl}{}^2v_{hkl}{}^2}}.\tag{2}$$

The quantity $\Delta_l$ is evidently the limiting value of the angular displacement in radians between the intensity maxima of the striations as $\Delta$ increases. Table I lists the measured values of the displacements (in radians) between maxima. Taking $\Delta_l=6.7\pm0.5\times10^{-4}$, and $\Delta_0-\Delta_1=13.2\pm0.2\times10^{-4}$, one calculates that $D=1400\pm100$A and $v_{420}=7.7\pm2.0\times10^{15}$ cm$^{-2}$. The value calculated from the definition of $v_{hkl}$ is $7.85\times10^{15}$ cm$^{-2}$. Using this value and the value 1400A for the thickness of the crystal, the calculated displacements between maxima of Table I are obtained. Evidently the greatest distance is between the central maximum and the next one; all the other distances are rather close to the limiting value. For a strong reflection the method provides a reasonably accurate determination of the structure factor for a reflection; for a weak reflection the percentage accuracy would be very low.

In the particular example cited here, no electron micrograph was obtained to substantiate the assumption of uniform thickness of the crystal. The agreement of the results with theory, however, supports this assumption.

We may now consider why in some cases cross-grating patterns, in which the Bragg law of incidence is apparently not obeyed, were observed and in other cases, Kikuchi-line patterns, which depend on those electrons incident at or very close to the Bragg angle being much more strongly diffracted than those incident at other angles. We first note that there is no logical contradiction between the occurrence of a strong reflection of rays incident at the Bragg angle, and weak reflections of rays incident at other angles for a particular crystal. Figure 5 is an example of a two-dimensional grating pattern and a Kikuchi-line pattern obtained from the same böhmite crystal; from the pattern of Fig. 8 for BeO we see that there is a 110 reflection of the entire aperture of illumination and displaced from that a broad Kikuchi line because of electrons inelastically scattered onto a direction making the Bragg angle with the 110 plane.

It is beyond the scope of the present paper to attempt to assess exactly the effect of disorder in a crystal on its diffraction pattern. In general the effect of such disorder will be to increase the intensity of diffraction of rays not incident at the Bragg angle.

TABLE II. "Rocking crystal" curve for MgO, 200 reflection.

| | 0 | $10^{-3}$ | $5 \times 10^{-3}$ | $10^{-2}$ | $1.285 \times 10^{-2}$ | $5 \times 10^{-2}$ |
|---|---|---|---|---|---|---|
| $I/I_0$ | 1/2 | 0.485 | 0.29 | 0.127 | 0.085 | 0.068 |
| $D_0$ (period of $\sin^2$ term, $A$) | 360 | 354 | 274 | 254 | 181 | 149 |

The dynamic theory of electron diffraction, however, does indicate that even for perfect crystals the kinds of diffraction pattern reported in this paper are to be expected. MacGillavry's formula (1) for the intensity of a reflection gives for the half-width of incidence of a reflection (neglecting the $\sin^2$ term)

$$\Delta_{\frac{1}{2}} \sim v_{hkl} d_{hkl},$$

$$\Delta_{\frac{1}{2}} \sim (\textstyle\sum_j)_{hkl} d_{hkl}^3,$$

so that wide angle reflections give the best Kikuchi lines, whereas the inner reflections give broad Kikuchi lines and readily contribute to cross-grating patterns. This relation is evident in the 110 and 220 Kikuchi lines of BeO in Fig. 8.

As an example of the intensity of diffraction for angles of incidence that are considerably different from the Bragg angle, it is of interest to apply MacGillavry's formula to calculate a "rocking curve" for the 200 reflection of MgO, using $\lambda = 5.4 \times 10^{-10}$ cm. In this calculation (Table II), the average value of the $\sin^2$ term is taken as ½.

Even at normal incidence ($\Delta = \theta_B$) the diffracted intensity is great enough to give noticeable reflection. One similarly computes for the intensities of the 220 and 400 reflections at $\Delta = \theta_{220}$ and $\Delta = \theta_{400}$ the values 0.011 and 0.0013, respectively.

Miss MacGillavry's formula (1) considers only the interaction between one diffracted ray and the incident ray, the treatment of the general case of the interaction of all the diffracted rays being very complicated. It is of interest, however, to present calculations for the symmetrical case of normal incidence on a cube face of MgO considering the interaction of all four of the 200 reflections and all four of the 220 reflections. This is a simple computation and serves to indicate if formula (1) is seriously wrong or if it gives results of the same order of magnitude as those obtained by more elaborate calculations. Only an outline of the methods of calculation and the results need be presented.[21]

Bethe's dynamical equations for this case are:

$$(K^2 - k_{000}^2)\psi_{000} + 4v_{200}\psi_{200} + 4v_{220}\psi_{220} = 0,$$

$$v_{200}\psi_{000} + \left(K^2 - k_{000}^2 - \frac{4\pi^2}{d_{200}^2} + v_{400} + 2v_{220}\right)\psi_{200} + (2v_{200} + 2v_{420})\psi_{220} = 0,$$

$$v_{220}\psi_{000} + (2v_{200} + 2v_{420})\psi_{200} + \left(K^2 - k_{000}^2 - \frac{4\pi}{d_{220}^2} + 2v_{400} + v_{440}\right)\psi_{220} = 0.$$

In these relations, $k_{000}$ is the magnitude of the wave vector of the undiffracted wave in the crystal; its three possible values are determined by equating the determinant of the linear equations to zero. The $\psi$'s are the amplitudes of the various waves in the crystal. With numerical coefficients inserted, the cubic characteristic equation is solved, and the relative amplitudes of the internal waves determined for each solution. The customary boundary conditions for the Laue case lead to the intensity formulae: ($D$ in Angstrom units)

$$I_{000} = 1 - 0.904 \sin^2(\pi D/146) - 0.032 \sin^2(\pi D/125) - 0.060 \sin^2(\pi D/67.4),$$

$$I_{200} = 0.192 \sin^2(\pi D/146) + 0.0088 \sin^2(\pi D/125) - 0.0083 \sin^2(\pi D/67.4),$$

$$I_{220} = 0.037 \sin^2(\pi D/146) - 0.0077 \sin^2(\pi D/125) + 0.0098 \sin^2(\pi D/67.4).$$

The average value for the intensity of the diffracted waves is $\bar{I}_{200} \cong 0.10$, $\bar{I}_{220} \cong 0.02$.

The precise numerical values obtained are of no great significance because of the very special case treated; what is important is that these values are of the same order of magnitude as those obtained from formula (1) so that this formula may be used to indicate reliably the extent of a cross-grating pattern for a perfect crystal.

An outstanding deficiency of the dynamic theory as presented is that it neglects the effect of incoherent scattering on the propagation of a

---

[21] See references 15 and 18 for a complete discussion.

wave system in a crystal. The energy of the electron beam in the crystal is pictured as oscillating back and forth between the incident and the diffracted beams without damping. The effect of incoherent scattering will be to increase the general background, to tend to destroy the diffracted rays—particularly for rays not incident at the Bragg angle,—and to cause the development of Kikuchi lines.

## V. CONCLUSIONS

One can conclude then that two-dimensional grating patterns are to be expected for crystals oriented with an important lattice vector approximately parallel to the electron beam, provided the crystals are not too thick and independent of whether or not the crystals are slightly disordered or warped. Kikuchi-line patterns are to be expected for most other crystals that are not excessively thick or disordered; the wide angle reflections will give the sharpest and best Kikuchi lines. Because of the shape of the "rocking crystal" reflection curves, crystals which give rise to two-dimensional grating patterns may also give rise to Kikuchi-line patterns.

## ACKNOWLEDGMENTS

The authors wish to thank E. G. Ramberg for reading the manuscript and checking some of the calculations.

---

## Calendar of Meetings

**June**

| | |
|---|---|
| 1–6 | American Society of X-Ray Technicians, Buffalo, New York |
| 1–6 | Society of Automotive Engineers, French Lick, Indiana |
| 8–10 | American Society of Refrigerating Engineers, Los Angeles, California |
| 9–14 | Symposium on Molecular Structure and Microscopy, Ohio State University, Columbus, Ohio |
| 12–14 | Canadian Association of Professional Physicists, University of Western Ontario, London, Ontario, Canada |
| 12–14 | National Organic Chemistry Symposium, Boston, Massachusetts |
| 16–19 | American Society of Mechanical Engineers, Chicago, Illinois |
| 16–20 | American Society for Testing Materials, Atlantic City, New Jersey |
| 16–20 | American Association for the Advancement of Science (Pacific Division), San Diego, California |
| 18–20 | American Association of Physics Teachers, Regional Meeting, University of Minnesota, Minneapolis, Minnesota |
| 18–21 | American Society for Engineering Education, Minneapolis, Minnesota |
| 19–21 | American Physical Society, Montreal, Quebec, Canada |
| 23–25 | American Society for X-Ray and Electron Diffraction, Montreal, Quebec, Canada |
| 23–25 | American Chemical Society (Mid-West Regional Meeting), Kansas City, Missouri |

**July**

| | |
|---|---|
| 11–12 | American Physical Society, Stanford University, California |
| 16–19 | American Society of Civil Engineers, Duluth, Minnesota |

**August**

| | |
|---|---|
| 21–23 | Society of Automotive Engineers (Transportation and Maintenance Meeting), Los Angeles, California |
| 26–29 | American Institute of Electrical Engineers, San Diego, California |

**September**

| | |
|---|---|
| 1–2 | Mathematical Association of America, New Haven, Connecticut |
| 1–4 | American Society of Mechanical Engineers, Salt Lake City, Utah |
| 8–12 | Instrument Society of America, Chicago, Illinois |
| 15–19 | American Chemical Society, New York, New York |
| 15–19 | Illuminating Engineering Society, New Orleans, Louisiana |
| 17–18 | Society of Automotive Engineers, Milwaukee, Wisconsin |

**October**

| | |
|---|---|
| 2–4 | Society of Automotive Engineers, Los Angeles, California |
| 15–18 | American Society of Civil Engineers, New Orleans, Louisiana |
| 20–24 | American Society for Metals, Chicago, Illinois |
| 21–25 | American Chemical Society, San Francisco, California |
| 23–25 | Optical Society of America, Cincinnati, Ohio |

# Microsecond Measurement of the Phosphorescence of X-Ray Fluorescent Screens*

FITZ-HUGH MARSHALL
*Westinghouse Research Laboratory, East Pittsburgh, Pennsylvania*

(Received February 4, 1947)

Microsecond square-wave x-ray pulses, repeated a thousand times a second, were obtained by operating an x-ray tube directly from a radar pulser. The phosphorescent build-up and decay of light emitted by fluorescent materials when irradiated by the x-ray pulses was observed with a multiplier phototube connected to an oscilloscope. Oscillograms of $CaWO_4$ (radiographic intensifying screen) and $BaSO_4$ indicate simple exponential response, with decay to $1/e$th intensity in 6 and 0.8 microseconds, respectively. $MgWO_4$ and Patterson B (fluoroscopic) and D (photofluorographic) screens were found to complete their main response to microsecond excitation in 10 to 100 microseconds, but their decay is not exponential.

A SUSTAINED interest in the subject of phosphorescence during the last decade has led to extensive experimental work, though rarely with x-ray excitation, and to various theoretical proposals to account for the many cases in which the observed decay does not follow a simple exponential law as would be expected for a simple monomolecular process.[1–4] Attempts have been made to interpret these more complicated cases as composites of superimposed exponential curves, as hyperbolic decay associated with a bimolecular mechanism, or as due to emission delays by electron traps. The present experiment is not sufficiently comprehensive to test the validity of these theories, for it had as its immediate purpose only the determination of the speed of response of standard types of commercial fluorescent x-ray screens. But the experimental method employed has proved so simple and effective that others may wish to develop it for a more detailed treatment, either for the study of x-ray screens as such or for a more fundamental investigation of phosphorescence. The microsecond timing technique used, instead of the more common millisecond timing, is clearly necessary to resolve rapid phosphorescence effects, which should not be disregarded in a general study. The use of x-rays, instead of the more usual ultraviolet or cathode-ray excitation, not only provides new test conditions but also simplifies absorption corrections when they are necessary, as in hyperbolic decay: in general the x-ray correction is less, can be determined accurately and easily, and does not involve taking into account a large change of intensity within individual crystals.

The equipment used in this study of the short persistence phosphorescence of x-ray fluorescent screens is illustrated in the block diagram of Fig. 1. An x-ray tube was operated directly from a radar pulser. The pulser provided microsecond square-wave potential pulses at 10 to 30 kv, repeated a thousand times a second. The x-ray tube current during the pulse was as high as 500 ma, but the corresponding average current was well within the rating of the air-cooled fluoroscopic tube used. With a water- or oil-

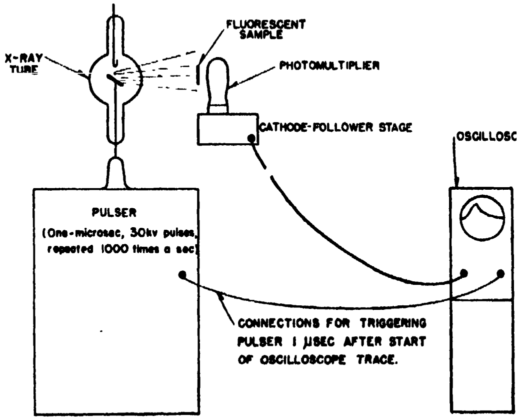

X-RAY TUBE
FLUORESCENT SAMPLE
PHOTOMULTIPLIER
CATHODE-FOLLOWER STAGE
OSCILLOSC
PULSER
(One-microsec, 30kv pulses, repeated 1000 times a sec)
CONNECTIONS FOR TRIGGERING PULSER 1 μSEC AFTER START OF OSCILLOSCOPE TRACE.

FIG. 1. Equipment for microsecond measurement of the phosphorescence of x-ray fluorescent screen materials.

---

* Reported at the Chicago meeting of the Am. Phys. Soc., June 22, 1946 (Phys. Rev. **70**, 114 (1946)).

[1] J. T. Randall, M. H. F. Wilkens, and G. F. J. Garlick, Proc. Roy. Soc. **A184**, 347–433 (1945).

[2] J. W. Strange and S. T. Henderson, Proc. Phys. Soc. London **58**, 369–401 (1946).

[3] R. T. Ellickson and W. L. Parker, Phys. Rev. **70**, 290–299 (1946).

[4] W. B. Nottingham, R. B. Nelson, and R. P. Johnson, J. App. Phys. **10**, 335–342 (1939).

cooled tube, much larger currents could have been obtained if needed, though probably with some reduction in the filament life because of operation at abnormally high temperatures. The x-rays irradiated the fluorescent sample, which for simplicity was taped directly against the R.C.A. multiplier phototube used to measure the light emitted, the phototube then being wrapped in black paper to exclude room light. The distance from the x-ray target to the fluorescent sample was about 15 cm. Stray high-frequency pickup from the pulser was practically eliminated by (1) covering the phototube with a grounded hood of aluminum foil to complete the system of standard grounded circuit shields, (2) introducing a cathode-follower stage mounted in pre-amplifier position directly under the phototube, and (3) grounding various extra points of the equipment for optimum results as determined by trial and error. The phototube and cathode-follower stage were mounted on top of the pulser along with the x-ray tube and its filament isolation transformer. The multiplier phototube was operated at about 90 v per stage from a d.c. supply, which was well filtered and well regulated.

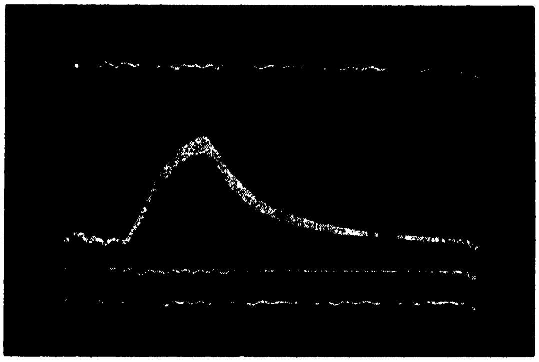


FIG. 2. Oscillogram of the growth and decay of the phosphorescent response of barium sulphate. The total time of the trace is 5 microseconds.

The cathode-follower stage was connected directly to the amplifier input of a DuMont 248 oscilloscope, no additional amplification being required. This oscilloscope, which has a number of special auxiliary circuit features, was used as a synchroscope to trigger the pulser a microsecond after the beginning of the rapid sweep of the cathode-ray beam. The sweep time could be varied and accurate time-marker intervals superimposed on the trace. From the phototube through the oscilloscope the response of the circuit was checked to be nearly flat to about 5 megacycles per second. The response of the phototube itself is good to about 100 megacycles per second.

The phosphorescent response of the fluorescent sample during and after the microsecond of irradiation by x-rays thus registered on the oscilloscope screen and was photographed. Additional exposures on the same film were used to introduce reference traces. Figure 2 is an oscillogram obtained in this way for barium sulphate, the fastest fluorescent material studied. The time represented by the complete curve is 5 microseconds. Trace *a* is the main response curve, trace *b* is a base line recorded with x-rays off and displaced downward to prevent confusion from overlapping of the small initial irregularities caused by stray pick-up, and traces *c* are similar lines above and below upon which microsecond time markers have been introduced for accurate time measurement.

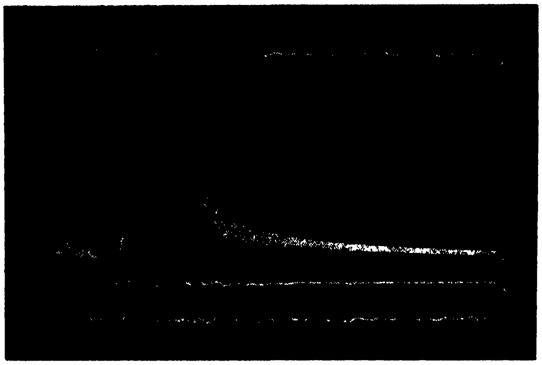


FIG. 3. Response to direct x-rays.

As will be shown, the response for barium sulphate is exponential. Following through the succession of events recorded in Fig. 2, we see that during the first microsecond there are no x-rays. Then the x-ray pulse starts, and the phosphorescent response begins to rise as the material is irradiated at constant intensity for a microsecond. During excitation active phosphorescent centers are created at a constant rate. But as the number of active centers increases,

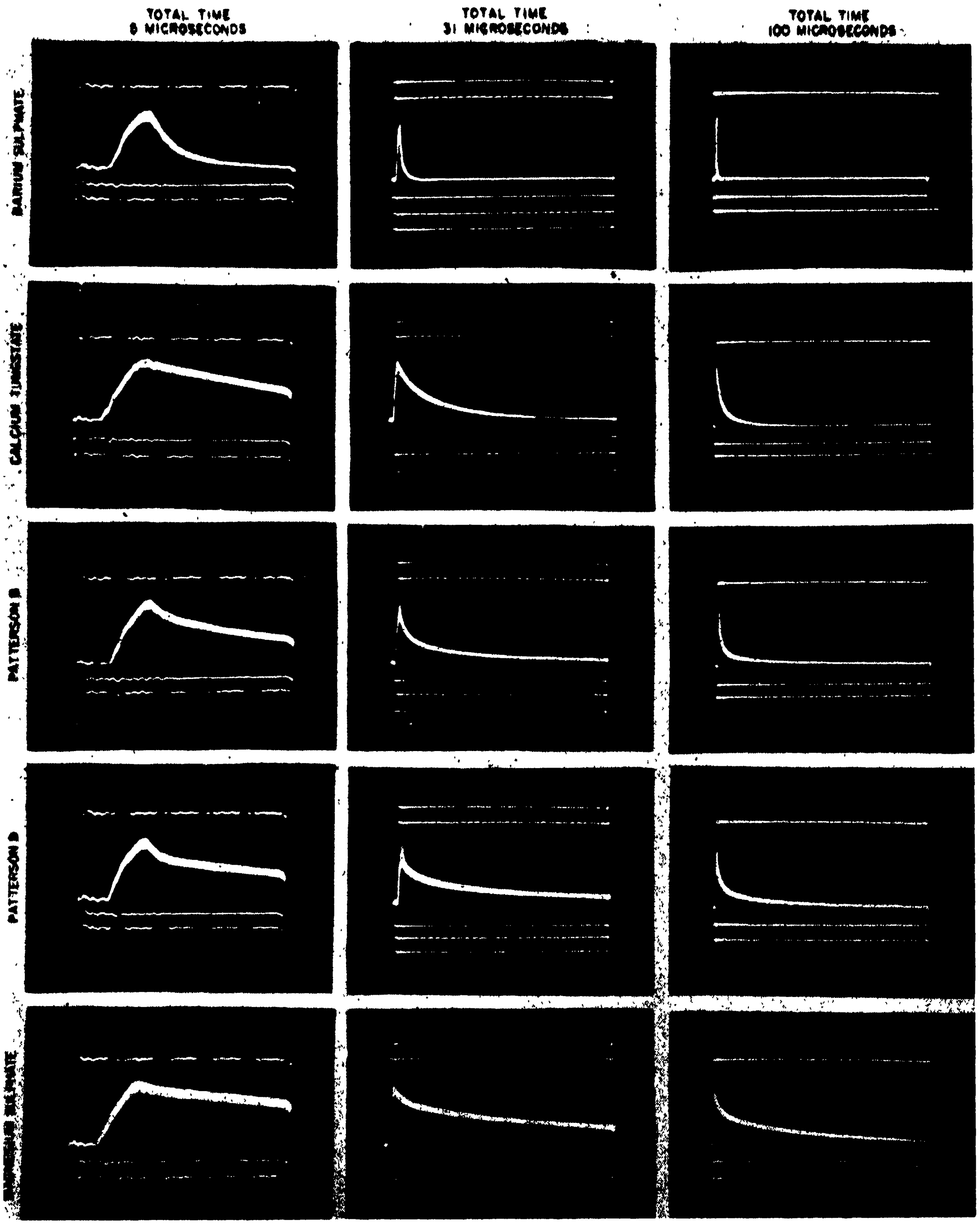


FIG. 4. Comparative oscillograms of the phosphorescent response of various fluorescent materials to a microsecond x-ray pulse.

so does the rate at which they disappear with the emission of light, this rate being always proportional to the number remaining. If the excitation were to continue indefinitely, the response would exponentially approach that equilibrium level *d* at which the active centers disappear as fast as they are created. But in this case the x-rays are turned off after a microsecond of irradiation, and the light intensity decays exponentially toward zero as the active centers disappear. For barium sulphate the phosphorescent period (time for decay to 1/*e*th intensity) is only 0.8 microsecond.

A direct test of the resolution and general effectiveness of the equipment is provided by Fig. 3. It so happens that the phototube responds weakly to direct x-rays with no intervening fluorescent screen. The oscillogram of Fig. 3 is for direct x-rays of extremely high intensity. The curve differs from a square wave because of the presence within the phototube of a weakly fluorescent material which phosphoresces exponentially with a period of 1.2 microseconds. When this superimposed response, similar to the one in Fig. 2, is subtracted, a reasonably square wave remains, resolved to about 0.2 microsecond.

The broadening of the main trace in all oscillograms recorded has been studied carefully and found to be due to noise of the statistical fluctuations in the number of x-ray quanta that are effective in producing phototube response.[5] The noise is exceptionally high in Fig. 3 for the following reason: although very few x-ray quanta are absorbed at the photosurface, each quantum which is absorbed produces a shower of photo-electrons.

In Fig. 4 is presented a complete set of oscillograms for the various fluorescent materials studied. The first, at the upper left, is a repetition of the 5-microsecond curve for barium sulphate shown in Fig. 2. On the 31-microsecond oscillograms, 10-microsecond time markers have been introduced in addition to the 1-microsecond markers. On the 100-microsecond oscillograms only 10-microsecond markers are included. This set of oscillograms covers the fastest and slowest materials studied and the three most commonly used types of x-ray screens. The x-ray intensity has been adjusted to bring the maximum response to approximately the same value for all curves.

The barium sulphate (which included 5 percent of lead sulphate) and the magnesium tungstate were selected from an assortment of available experimental powders which had been prepared for an unrelated purpose in another laboratory; the fluorescent intensity of each of these powders was very low in comparison with that of commercial x-ray screens. The calcium tungstate, which fluoresces blue, was a standard commercial radiographic intensifying screen; the shape of the response curve appeared to be the same for the several brands, grades, and thicknesses of screens tested, though this has not been checked by careful measurement. The Patterson B screen fluoresces with the familiar yellow-green of fluoroscopic screens, of which it is a standard example; it may be classed as a zinc-cadmium sulphide screen. The Patterson D screen, which fluoresces blue, is one which has been developed recently for photofluorography; it may be broadly classed as a zinc sulphide screen.

The simplest kind of phosphorescent response is an exponential one, the equations for the intensity, $I_g$ during excitation or growth and $I_d$

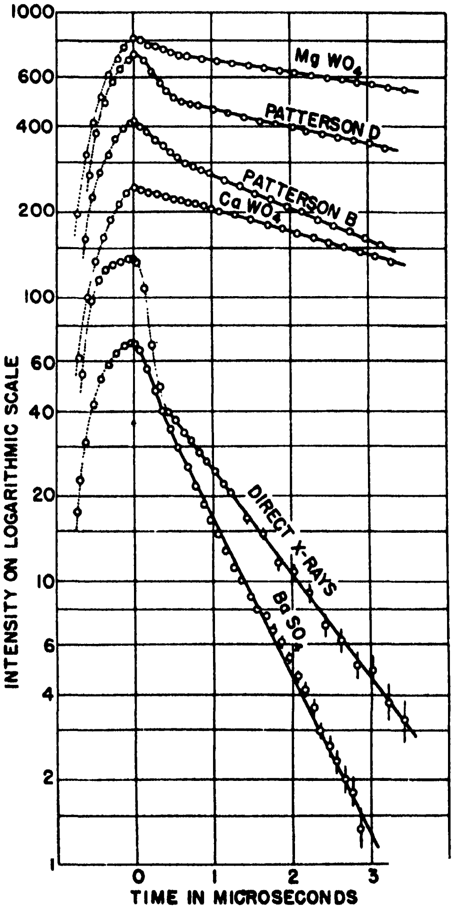


FIG. 5. Response curves plotted from measurement of 5-microsecond oscillograms.

---

[5] Fitz-Hugh Marshall, J. W. Coltman, and L. P. Hunter, Rev. Sci. Inst. tentatively scheduled for publication in July, 1947 issue.

during decay, being of the form

$$I_g = A(1-e^{-t/T}) \quad \text{and} \quad I_d = Be^{-t/T},$$

where $t$ is time, $T$ is the phosphorescent period, and $A$ and $B$ are constants. While there is no particular reason for expecting exponential response, it is a simple matter to check whether a response is exponential from either the growth or the decay portions of the curve. In this experiment the decay portion yielded much greater accuracy than the steep, poorly resolved growth portion.

In Figs. 5 and 6 the decay curves for the various materials studied are plotted on semi-logarithmic paper, on which a simple exponential decay would appear as a straight line. The data are taken from careful measurement of the oscillograms. Intensities have been multiplied by arbitrary factors to separate the curves conveniently. An attempt has been made to adjust the time scale so that the decay is shown starting at zero time. In Fig. 5 portions of the earlier growth curves are included as dotted lines for reference; the initial drop-off of the curve for direct x-rays is also dotted, for it represents the poorly resolved response of the circuit to the cut-off of the square wave rather than a phosphorescent decay. The spread shown for lower points on some of the curves indicates error due to uncertainty in measurement of small ordinates on the oscillograms; the deviation shown is about a thirtieth of the width of the recorded trace. Somewhat larger errors may very well be present from non-linearity in registration of the cathode ray tube.

Within the limits of experimental error, the response appears to be exponential for calcium tungstate, barium sulphate, and the fluorescence of the phototube under direct x-rays, the phosphorescent periods being 6, 0.8, and 1.2 microseconds, respectively. The several small but clearly discernible deviations from linearity in the curves for those materials can be explained qualitatively in the following way. A slight bend for barium sulphate during the first microsecond of decay is attributable to a small component of direct x-ray action on the phototube. A bend in the calcium tungstate curve at 5 microseconds must be ascribed to an overswing in the amplifying circuit, for at the corresponding time a dip through a minimum is clearly observed on the 31- and 100-microsecond oscillograms for barium sulphate and also on the oscillograms in Fig. 7. As for the deviation of the points from a straight line at the lower end of the calcium tungstate curve, other experiments[1, 4] indicate that calcium tungstate is so rapid in its decay that it would be uncalled for to suspect a progressive upswing in the curve on the basis of these few points of large experimental error; however, Strange has reported two approximately exponential cathodoluminescent periods for calcium tungstate.[2]

A detailed analysis has not been attempted for the magnesium tungstate and sulphide screens. Their decays are obviously not exponential. The comparatively sharp bends near the start of these curves are evidently real, and they suggest the presence of superimposed decay curves. The later portions of these curves fit simple hyperbolic decay a little better than exponential decay, but the difference is not considered significant in the absence of additional supporting evidence and a more complete knowledge of the composition of the commercial screens involved. No change in the shape of the curves was noticed on the oscilloscope with variation of x-ray potential or intensity, as might be expected for hyperbolic decay. If a more detailed experiment is undertaken by others, this point should be checked more carefully, along with temperature dependence. Other studies which might be of interest include the response as a function of wave-length

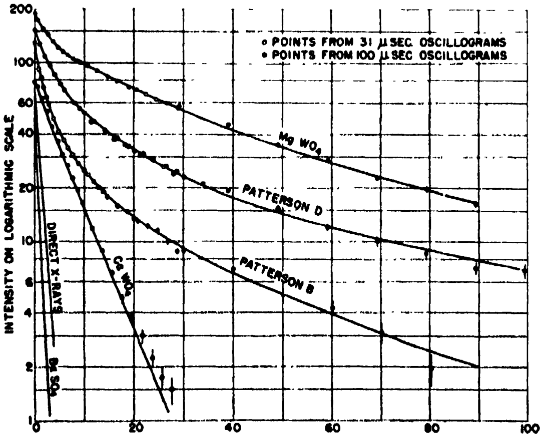


FIG. 6. Response curves extended through 100-microseconds of decay.

of light emitted by the screen and the effect of infra-red stimulation of the screen.[3]

In interpreting the oscillograms and curves in this paper, it is important to note that the unusual technique of microsecond excitation has altered the shape of all curves which are not simple exponential responses in such a way as to accentuate rapid phosphorescence effects and to suppress slow persistence effects. Thus the results may not immediately appear to be consistent with observations made under more conventional conditions. This was forcibly brought out during the experiment by an attempt to observe the phosphorescence of willemite, a screen with a persistence comparable to the flicker time of the eye. Although the willemite could be seen to glow brightly when irradiated with x-rays, an oscillogram obtained according to the normal procedure of this experiment (i.e., one made with x-ray intensity adjusted to bring the peak response to the usual amplitude) revealed only the response of the phototube to direct x-rays and was indistinguishable from an oscillogram for direct x-rays without the intervening willemite. A difference was finally obtained, as illustrated in Fig. 7, by greatly increasing the x-ray intensity (and the amplification), past the point of overloading the amplifier at the peak. In the second part of Fig. 7, the main evidence of the luminescence of the willemite is the broadening of the curve by the high noise level of the light, which is intense but which does not decrease appreciably during the short interval between pulses.

Let us assume for simplicity that the reciprocity law holds perfectly for each of the components involved in the oscillogram at the right in Fig. 7, that is, that the total area under the curve (extended to infinite time) associated with each component is independent of the time of x-ray excitation so long as the total quantity of x-rays remains unchanged. It is evident that if the same quantity of x-rays had been introduced during a pulse time of 10 microseconds, instead of one microsecond, the willemite portion of the response would not have been appreciably altered, but the direct x-ray portion of the response would have been spread over approximately ten times as much time and would therefore have risen only a tenth as high above the willemite response. With a pulse time of a thousandth of a second, such as might ordinarily be used for studying willemite, it is doubtful that any direct x-ray component could be detected, even with the x-ray intensity increased further to bring the peak response to the usual level.

FIG. 7. Comparative 100-microsecond oscillograms for direct x-rays and willemite. The amplifier gain and the x-ray intensity have been so greatly increased as to cause overloading at the peaks.

This dependence of the shape of the response curve upon the length of time of excitation might provide the basis of a method for isolating component decays in composite curves. A pulser with variable pulse time would permit observation of the change of shape of the response curve with pulse time. A further useful innovation might be that of Strange of balancing out the main part of the response with counter exponential condenser discharge curves and measuring only the difference.[2] It should be noted that hyperbolic decay curves would behave differently from composite exponential curves, there being an actual energy shift from long-persistence effects to short-persistence effects as the pulse length is shortened.

It should be mentioned that in the present experiment all persistence after a thousandth of a second is rendered completely indistinguishable by the beginning of the next sweep cycle and the advent of a new x-ray pulse; the zero ordinate point seen at the beginning of each oscillogram trace is merely the response level at the end of the previous millisecond cycle.

Let us finally consider the practical importance of the results of this experiment with respect to the speed of response of commercial x-ray screens. It is necessary to introduce supplementary information from other sources, inasmuch as long-persistence effects of relatively high energy might be present without detection by the present experiment.

Under ordinary conditions calcium tungstate does not show any detectible persistence. When a calcium tungstate screen is whirled as a disk before a narrow beam of x-rays, there is no noticeable blurring of the edge of the primary spot on the screen covered by the x-ray beam. We may therefore conclude that the curves for calcium tungstate recorded in this experiment include all of the response which is of practical importance and that the response is almost complete in $10^{-5}$ second.

In contrast, the sulphide screens, especially Patterson D, are commonly known to have long-persistence effects which are visible for some seconds or minutes after exposure to intense x-rays. Since these are not detected in the present experiment, it is necessary to consider whether a large fraction of the total light emitted by the screen is associated with the long-persistence effects. The case for Patterson D is of greatest practical importance, for it is usually found to be the brightest screen as measured photographically or photoelectrically. A rotating-disk test shows the primary spot on the screen to be bright and sharply delineated from the phosphorescent trail. On the basis of this and other tests, it appears safe to say that the main useful part of the response (perhaps three-fourths for Patterson D, more for Patterson B) is included in the oscillograms of this experiment and is complete in 20 microseconds or so. Thus the amount of long-persistence light is too small to be especially important from the standpoint of efficiency; on the other hand, it is large enough to be a disturbing factor in fast response applications for which it is essential that there be no detectible residual from earlier x-ray events.

As a practical case in point, we may consider x-ray motion pictures. These are now generally obtained by photographing the fluorescent screen. When Patterson D is used with conventional x-ray and camera equipment, noticeable blurring or multiple-image effects occur, especially at the leading edge of a dark area moving into a bright area. This blurring does not occur when calcium tungstate screen is used, but the effective intensity is less.

These conclusions also apply to the photomultiplier x-ray detector.[5] This detector consists of a fluorescent screen wrapped around a multiplier phototube and is therefore equivalent to the x-ray detector used in this experiment. Speed of response is important in the analysis of fast x-ray phenomena. Here again the rapid response and high intensity of Patterson D make it very suitable for many applications. However, the long-persistence effects of this screen cannot always be disregarded. For example, when the detector is used as an integrating exposure meter to terminate a radiographic exposure after the incidence of a given quantity of x-rays, the long-persistence effects result in an objectionable dependency upon exposure time and upon recent exposure history.

# X-Ray Diffraction Studies of Chrome-Steel Slags[1]

G. P. Chatterjee and S. S. Sidhu
*University of Pittsburgh, Pittsburgh, Pennsylvania*

(Received March 21, 1947)

X-ray diffraction studies of the crystalline constituents in the acid open hearth chrome-steel samples showed that chromium exists in these samples as a chrome-iron spinel of the form, $FeO \cdot Cr_2O_3$. Its density was 5.109 g/cc and its melting 2160°C. It crystallized in the face-centered cubic structure with $a_0 = 8 \cdot 348$A and $Z=8$.

## I. INTRODUCTION

THE essential major elements in most of the common alloy steels are chromium (Cr), nickel (Ni), and manganese (Mn). Nickel is not oxidized during steel making. Of the other two elements, Cr is more costly, and its loss in slag through oxidation or any other combination increases the cost of steel to a greater extent than an equivalent loss of Mn. The nature of control methods, which may lead to decreased losses of Cr, depends on the knowledge of the different phases in which Cr is present in the slag. The object of this investigation was to study by x-ray diffraction methods if Cr exists in the acid open hearth chrome-steel slag as (a) free $Cr_2O_3$ or CrO, (b) solute in FeO solvent, (c) silicates of chromium, and/or (d) spinels of the type $FeO \cdot Cr_2O_3$ or $CrO \cdot Fe_2O_3$.

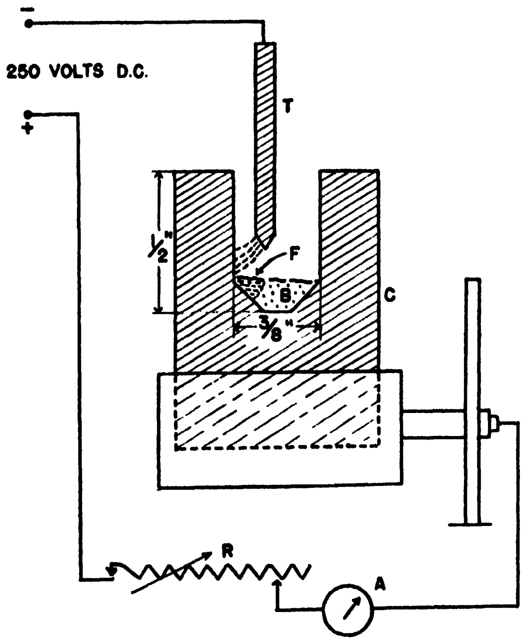


Fig. 1. The schemetic diagram of the small indirect arc furnace for preparing synthetic spinel. $T$ = top electrode; $C$ = rotating crucible; $F$ = fused charge; $B$ = unfused charge; $R$ = variable resistance; and $A$ = ammeter.

## II. EXPERIMENTAL PROCEDURE

Acid open hearth chrome-steel slag samples were obtained from regular heats of steel and poured into Fluidimeter.[2] The slag solidified and cooled rapidly inside the mould to preserve nearly the original conditions in which the different constituents existed in the molten slag. From the chemical analyses of these samples, samples of synthetic slag were prepared. A mixture of pure FeO and $Cr_2O_3$ (mole proportion by weight) was put in graphite crucible $C$ of a small 250-volt d.c. arc furnace as shown in Fig. 1. A small amount of borax was added to facilitate rapid melting. The top electrode $T$ was placed slightly eccentric with respect to the bore of the crucible. An indirect arc was struck on the side of the crucible while the crucible was rotated. In one complete rotation of the crucible, the entire mixture was melted beneath the arc. The time of rotation was 20 seconds and the current through the electrodes was 10 amperes. During the arcing the electrode $T$ was kept negative and the crucible positive. It was found by optical pyrometer that the positive electrode was always at higher temperature than the negative electrode.

---

[1] The paper was presented, in part, at the joint meeting of the American Society for X-Ray and Electron Diffraction and the Electron Microscope Society of America, Pittsburgh, Pennsylvania (Dec. 1946).

[2] Fluidimeter designed by the Acid Open Hearth Research Association under the direction of Dr. G. R. Fitterer, University of Pittsburgh, Pittsburgh, Pennsylvania.

To remove borates and other contaminants formed during melting, the synthetic samples were powdered and boiled in 1:2 HCl solution for a few minutes in a test tube. The solution with the residue was filtered and washed several times with hot water until the filtrate was free from iron salts and chlorides. The residue was dried in a muffle furnace at 120°C for one hour.

X-ray diffraction patterns of both the natural and synthetic slag samples were made by the usual powder method with filtered Mo$K\alpha$ radiation and a cylindrical camera of 171.9 mm effective diameter.

TABLE I. Powder diffraction data.

$d$ = interplanar spacing in A units, $I/I_0$ = estimated relative intensity. V = very; S = strong; M = medium; W = weak.

| Chrome steel slag | | Synthetic slag | | Faylite | |
|---|---|---|---|---|---|
| $I/I_0$ | $d$ | $I/I_0$ | $d$ | $I/I_0$ | $d$ |
| M | 4.82 | M | 4.83 | — | — |
| M | 3.70 | — | — | S | 3.71 |
| — | — | — | — | M | 3.53 |
| S(broad) | 2.93 | M | 2.95 | S | 2.85 |
| V.S. | 2.54 | V.S. | 2.51 | M | 2.55 |
| M | — | — | | M | 2.29 |
| M(broad) | 2.08 | M | 2.08 | W | 2.06 |
| S | 1.91 | S | 1.91 | V.W. | 1.91 |
| M | 1.72 | W | 1.71 | S | 1.76 |
| S | 1.61 | S | 1.61 | M | 1.61 |
| W | 1.54 | — | — | M | 1.51 |
| S | 1.492 | S | 1.488 | — | — |
| W | 1.442 | — | — | M | 1.42 |
| W | 1.352 | V.W. | 1.328 | W | 1.33 |
| M | 1.281 | M | 1.279 | V.W. | 1.277 |
| W | 1.216 | W | 1.210 | — | — |
| — | — | — | | V.W. | 1.197 |
| W | 1.176 | V.W. | 1.170 | V.W. | 1.176 |
| M | 1.111 | V.W. | 1.115 | — | — |
| W | 1.079 | M | 1.083 | — | — |
| — | — | W | 1.045 | — | — |
| M | 0.968 | M | 0.965 | — | — |
| — | — | W | 0.933 | — | — |
| W | 0.878 | W | 0.878 | — | — |
| M | 0.858 | M | 0.854 | — | — |

### III. DISCUSSION OF RESULTS

The x-ray diffraction data obtained from the powder patterns are given in Table I. It is apparent from the data that the chrome-steel slag contains the same compound or compounds that are present in the synthetic slag and also small amounts of faylite, $(FeO)_2SiO_2$. The analysis of the diffraction data of the synthetic slag sample shows that it contains no free FeO, $Cr_2O_3$, nor a solid solution of FeO and $Cr_2O_3$, but a compound formed from the two oxides. That this compound is a spinel similar to chromite ($FeO\cdot Cr_2O_3$), one of the natural spinels reported by Dana[3] may be shown from its crystal structure as determined below:

For cubic crystal system

$$d_{hkl}=a_0/(h^2+k^2+l^2)^{\frac{1}{2}},$$

or

$$(h^2+k^2+l^2)=a_0^2\times\frac{1}{d_{hkl}^2}, \tag{1}$$

where $d_{hkl}$ = interplanar spacing corresponding to plane $hkl$ and $a_0$ = side of the unit cubic cell. For face-centered cubic crystal, the Miller indices of the planes that given reflections and the sum of the squares of the Miller indices are given in Table II. By plotting $(h^2+k^2+l^2)$ as function of $1/d^2$, one gets a straight line as shown in Fig. 2. The slope of this line as given in Eq. (1) is $a_0^2$. An accurate value of $a_0$ thus determined is 8.348A.

The density of the synthetic slag sample determined by the usual picnometer method was 5.109 g/cc. The molecular weight of the compound $FeO\cdot Cr_2O_3$ is 223.87 g/mole. Therefore the number of molecules per unit cell of the compound is

TABLE II. Interplanar spacings and the corresponding Miller indices of the first eight lines of the synthetic spinel.

| $I/I_0$ | $d$ in A | $1/d^2$ in $A^{-2}$ | Miller indices $hkl$ | $(h^2+k^2+l^2)$ |
|---|---|---|---|---|
| M | 4.83 | .043 | 111 | 3 |
| M | 2.95 | .115 | 110(2) | 8 |
| V.S. | 2.51 | .158 | 113 | 11 |
| M | 2.08 | .231 | 100(4) | 16 |
| S | 1.91 | .274 | 133 | 19 |
| W | 1.71 | .342 | 112(2) | 24 |
| S | 1.61 | .386 | 111(3), 115 | 27 |
| S | 1.48 | .457 | 110(4) | 32 |

$$Z=\frac{5.109\text{ g/cc}\times(8.348)^3\times10^{-24}\text{ cc}\times6.023\times10^{23}\text{/mole}}{223.87\text{ g/mole}}=8.00.$$

---

[3] E. S. Dana, *A Text-Book of Mineralogy* (John Wiley and Sons, New York), p. 418.

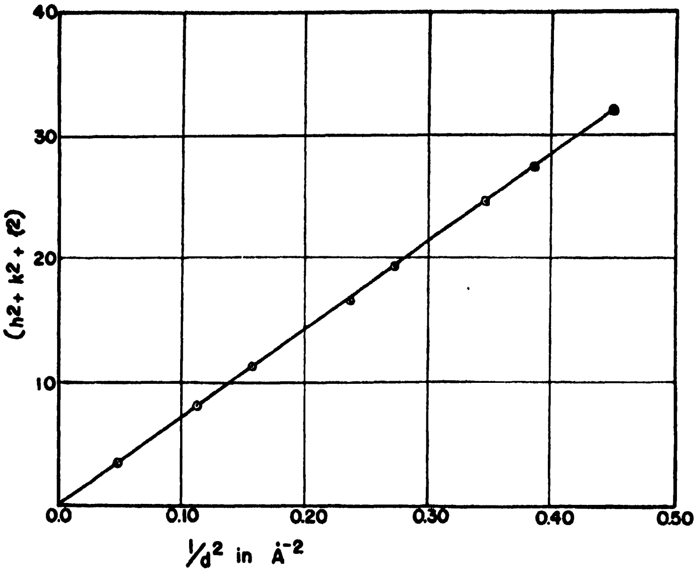


FIG. 2. $(h^2+k^2+l^2)$ as a function of $1/d^2$ for $FeO \cdot Cr_2O_3$.

If, on the other hand, it is assumed that the compound formed is $CrO \cdot Fe_2O_3$ whose molecular weight is 227.71 g/mole the number of molecules per unit cell is

$$Z=\frac{5.109\text{ g/cc}\times(8.348)^3\times10^{-24}\text{ cc}\times6.023\times10^{23}\text{/mole}}{227.71\text{ g/mole}}=7.86.$$

Since for a compound the number of molecules per unit cell must necessarily be whole number, it is evident that the chemical formula of the compound formed is $FeO \cdot Cr_2O_3$ with $Fe^{++}$ as the divalent element and $Cr^{+++}$ as the trivalent element.

## IV. CONCLUSIONS

1. The crystalline constituents in the acid open hearth chrome-steel slag samples are (a) chrome-iron spinel of the type $FeO \cdot Cr_2O_3$[4] and (b) small amounts of $(FeO)_2SiO_2$.

2. The density of the spinel as determined by the picnometer method is 5.109 g/cc and its melting point about 2160°C.

3. The spinel crystallizes in the face-centered cubic system with lattice constant $a_0$ as 8.348A and the number of molecules per unit cell is 8.

## ACKNOWLEDGMENT

The authors express their thanks to G. R. Fitterer for his very valuable suggestions and his assistance in supplying some of the necessary materials for this investigation.

---

[4] More exactly $(FeO, MnO) \cdot Cr_2O_3$, since the acid open hearth slag sample contains some amount of MnO which is usually present as $(MnO)_2SiO_2$ dissolved in $(FeO)_2SiO_2$. It is not unlikely, however, that some MnO may replace FeO in the spinel $FeO \cdot Cr_2O_3$. The atomic weight of Mn differs only by 0.92 from that of Fe and the atomic scattering factor of Mn does not differ much from that of Fe, the effects of replacing FeO to a small extent by MnO is therefore hardly detectable.

# Spherical Aberration of Compound Magnetic Lenses

L. MARTON* AND K. BOL
*Stanford University, Stanford University, California*

(Received January 15, 1947)

A reduction of the spherical aberration of strong electron lenses can be achieved by a strong lens as a virtual image former and by a transformation of the image in a real one by means of one or more weak lenses. Calculations are carried out for bell-shaped magnetic fields of the axial field distribution

$$H(z)=\frac{H_0}{1+(z/a)^2},$$

and numerical values of the achieved reduction of the aberration are given for different lens strength, magnification, and image distance of the compound system.

EARLY in the development of electron optics, it was recognized that spherical aberration is one of the most serious limiting factors in the performance of electron lenses. From the beginning, efforts have been directed toward a reduction of this defect. One of the earliest attempts was that of Scherzer[1] who demonstrated that for a thin and weak electrostatic lens an axial field distribution of the form $\phi(z)=\phi_0 e^{-Bz^2}$ gives the minimum spherical aberration. The first similar calculation for a magnetic lens was carried out by Glaser.[2] Later Recknagel[3] and Plass[4] contributed to the same subject. All these calculations deal with very weak lenses; the only practical lenses with somewhat reduced aberration were designed by Gray[5] and Law.[6]

Scherzer demonstrated in another paper[7] that the spherical aberration of electron lenses can never be reduced to zero. Rebsch and Schneider[8] showed furthermore that no combination of electric and/or magnetic lenses can have less aberration than any of the components alone would have. Already at that time, however, Rebsch[9] demonstrated that, theoretically, the aperture defect may be reduced below any prescribed limit if in a compound lens the physical size of the component lenses is reduced proportionately. For obvious reasons this method is not very practical.

A somewhat different approach to the problem of reduction of the spherical aberration of short focus, thick electron lenses was briefly reported eight years ago by one of us.[10] It consists of combining a strong lens, forming a virtual image, with one or more weak lenses which convert the virtual image into a real image. This procedure is universally adopted for the construction of light microscope objectives and other light optical systems, and a more thorough investigation of its application to electron optical systems seems desirable. In what follows, a quantitative evaluation of the reduction of the spherical aberration of strong magnetic lenses is given. For this purpose the theory of strong magnetic lenses, developed by Glaser,[11] must be somewhat extended. Since this discussion is limited to the reduction of spherical aberration, the reduction of other aberrations by the use of compound lenses, which was recently reported,[12] will be completely omitted.

Glaser has shown[11] that for practical fields both the spherical and chromatic aberrations have minima when the object is near the center of the lens, for which position the image is virtual for practically all lenses.

---

[*] Now at the National Bureau of Standards, Washington 25, D. C.

[1] O. Scherzer, Zeits. f. Physik **101**, 23 (1936).

[2] W. Glaser, Zeits. f. Physik **109**, 700 (1938); **116**, 19 (1940).

[3] A. Recknagel, Zeits. f. Physik **117**, 67 (1940).

[4] G. N. Plass, J. App. Phys. **13**, 49 (1942); **13**, 542 (1942).

[5] F. Gray, Bell Sys. Tech. J. **18**, 1 (1939).

[6] R. R. Law, Proc. I.R.E. **25**, 954 (1937).

[7] O. Scherzer, Zeits. f. Physik **101**, 593 (1936).

[8] R. Rebsch, and W. Schneider, Zeits. f. Physik **107**, 138 (1937).

[9] R. Rebsch, Ann. d. Physik **31**, 551 (1938).

[10] L. Marton, Phys. Rev. **55**, 672 (1939). (Abstract: more detail is given in U. S. Patent No. 2,233,264.)

[11] W. Glaser, Zeits. f. Physik **117**, 285 (1941).

[12] J. Hillier, J. App. Phys. **17**, 411 (1946).

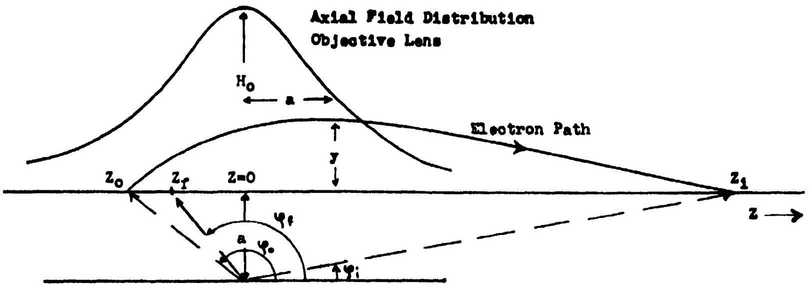


FIG. 1. Representation of symbols and coordinates used in Case I, a single objective lens.

Our problem, then, is to determine how small is the additional aberration introduced by the weak lenses. If it is sufficiently less than the decrease achieved by the formation of a virtual image by the first lens, our purpose is accomplished.

### BASIC THEORY FOR FIELDS OF FORM $H_0/[1+(z/a)^2]$

It was first shown by Glaser[1] that the spherical aberration produced by a magnetic lens with field distribution $H$ along the $z$-axis is given by:

$$\frac{\Delta r}{M}=\alpha^3\frac{e}{96mV}\int_{z_0}^{z_i}\left(\frac{2e}{mV}H^4+5H'^2-HH''\right)y_1^4dz. \tag{1}$$

Here $y_1$ represents the path equation of an electron satisfying the initial condition:

$$y_1(z_0)=0,\quad y_1'(z_0)=1. \tag{2}$$

As for the other symbols: $\Delta r$ is the divergence in the image plane, $z_i$, of a beam of aperture $\alpha$ at $z_0$, $M$ is the magnification, $V$ the electron energy in terms of the accelerating voltage, and $e/m$ the ratio of charge to mass of the electron. If $H$ is given by

$$H=H_0/[1+(z/a)^2], \tag{3}$$

then the path equation is

$$y=a(1+(z/a)^2)^{\frac{1}{2}}\{D_1\sin[(1+k^2)^{\frac{1}{2}}\operatorname{arccot}z/a]+D_2\cos[(1+k^2)^{\frac{1}{2}}\operatorname{arccot}z/a]\}, \tag{4}$$

$D_1$ and $D_2$ being arbitrary, while $k^2=eH_0^2a^2/8mV$. $a$ is the so-called half-width of the magnetic field. $k^2$ is a measure of the refracting power. The equation is put in a more palatable form if one sets $z/a=\cot\varphi$ (see Fig. 1) and $(1+k^2)^{\frac{1}{2}}=\omega$. One then obtains

$$y=aC_1\frac{\sin\omega(\varphi-\varphi_0)}{\sin\varphi}, \tag{5}$$

one constant having been chosen so that $y(z_0)=0$. For convenience we quote a few more formulas from Glaser:

$$z_{F_0}=a\cot(n\pi/\omega),\quad z_{F_i}=-a\cot(n\pi/\omega), \tag{6a}$$

$$f_0=\frac{a}{\sin(n\pi/\omega)},\quad f_i=\frac{-a}{\sin(n\pi/\omega)}, \tag{6b}$$

$$N<\omega<N+1, \tag{6c}$$

$N$ being the number of focal positions. Here $f$ is the focal length and $z_F$ is the location of the focal point.

The aberration for a single lens of this type has been thoroughly studied, whence we give just a resumé of some of Glaser's work.

Our problem, to be sure, is to study the aberration produced by a compound objective. This, however, we can immediately do by extending Eq. (1) over the whole system, as long as we obey the restriction that the radial distance of a paraxial ray be zero at the endpoints of the path of integration.

*Case I. Single lens.* The equation for the spherical aberration is obtained by integrating Eq. (1). From it follows that:

$$\frac{\Delta r}{M}=\alpha^3\frac{a_1k_1^2}{6\omega_1^4}\frac{1}{\sin^4\varphi_0}K_1, \tag{7}$$

where $K_1$ is given by

$$K_1=\int_{\varphi_0}^{\varphi_i}[2(4k_1^2-3)\sin^2\varphi+7]\sin^4\omega_1(\varphi-\varphi_0)d\varphi. \tag{8}$$

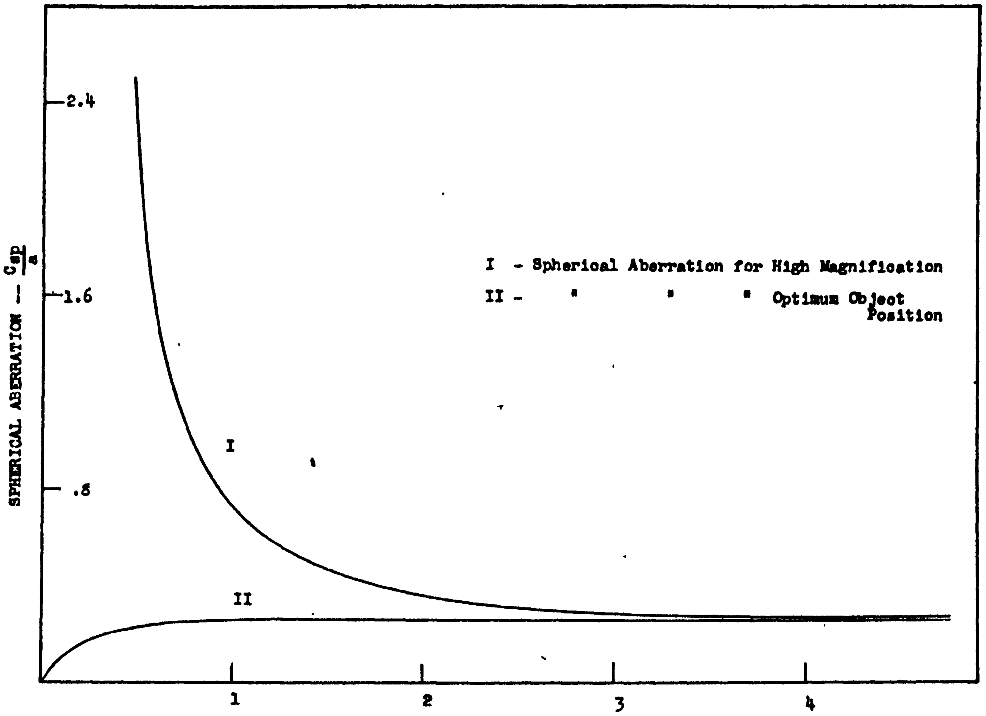


FIG. 2. Spherical aberration *vs.* lens strength for high magnification and optimum object position. (From Glaser.)

Since we need other upper limits later on, we leave it variable and obtain for the complete expression:

$$K_1(\varphi_0,\varphi)=\tfrac{3}{2}\omega_1^2(\varphi-\varphi_0)-\omega_1\sin 2\omega_1(\varphi-\varphi_0)+\tfrac{1}{8}\omega_1\sin 4\omega_1(\varphi-\varphi_0)-\frac{3(4k_1^2-3)}{16}\sin 2\varphi$$

$$+\frac{3}{4}\frac{\omega_1^4}{k_1^2}\frac{4k_1^2-3}{4k_1^2+3}\sin 2\varphi_0+\frac{4k_1^2-3}{4k_1^2}\omega_1\sin 2\omega_1(\varphi-\varphi_0)\cos 2\varphi-\frac{4k_1^2-3}{4k_1^2}\cos 2\omega_1(\varphi-\varphi_0)\sin 2\varphi$$

$$-\frac{1}{8}\frac{4k_1^2-3}{4k_1^2+3}\omega_1\sin 2\omega_1(\varphi-\varphi_0)\cos 2\varphi+\frac{1}{16}\frac{4k_1^2-3}{4k_1^2+3}\cos 4\omega_1(\varphi-\varphi_0)\sin 2\varphi. \quad (9)$$

For high magnification, however, $\varphi_0 \cong \varphi_F$ while $\varphi_i \sim 0$, and from the lens equation (6) $\omega\varphi_F=\pi$, so that the expression is much simplified. The aberration for $\varphi_0=\varphi_F$ as a function of $k^2$ is given by Fig. 2.

The aberration is, to be sure, not a minimum for $\varphi_0=\varphi_F$. The presence of the factor $\sin^{-4}\varphi_0$ suggests immediately that the object should be close to the center, or $\varphi_0 \sim \pi/2$, as was mentioned in the introduction. Indeed, Glaser also plots the aberration for the optimum object position, which we give in Fig. 2.

For high $k^2$ the aberration is not appreciably altered by picking the optimum position; for the lower $k^2$ that are actually used in practice $(1<k^2<2)$, however, improvement is possible.

There are two objections to having the object right at the optimum position: the magnification is almost unity for practical values of $k$, and the image is virtual.

In order to accomplish anything we must therefore sacrifice some improvement in the aberration to gain magnification, and we must put in at least a second lens to attain a real image. We must, therefore, invesigate next the aberration introduced by the second lens.

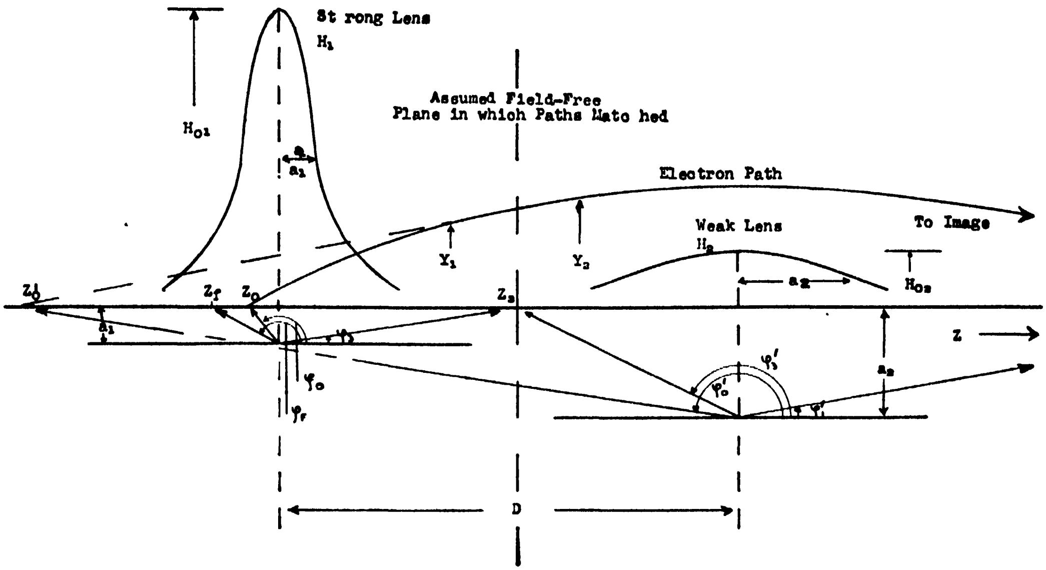


FIG 3. Representation of symbols and coordinates used in Case II, for a doubly compound objective lens.

*Case II. Two lenses.* We first define our axial field $H$ as follows:

$$H=\begin{cases} H_1=H_{01}/[1+(z/a_1)^2] \text{ for } -\infty<z<z_3, \\ H_2=H_{02}/\{1+[(z-D)/a_2]^2\} \text{ for } z_3<z<\infty. \end{cases} \quad (10)$$

$D$ is the distance between lens centers.

Quite arbitrarily we choose $z_3$ so that

$$H_1(z_3)\cong H_2(z_3).$$

We also assume that the two fields do not interact to an appreciable extent. This is done so that we can use the path equations of the previous section for each field, the constants of the second being adjusted to match the first at $z_3$. Only in the absence of such interaction are the calculations strictly valid.[11, 12]

Equation (1) then becomes

$$\frac{\Delta r}{M}=\alpha^3\frac{e}{96mV}\left[\int_{z_0}^{z_3} f(H_1)y_1^4dz+\int_{z_3}^{z_i} f(H_2)y_2^4dz\right]. \quad (11)$$

The integrals can be transformed into that of Eq. (8) by adopting a new coordinate system for the second lens. We take $z'=z-D$ and $z'/a_2=\cot\varphi'$ (Fig. 3). Applying furthermore the initial condition (2) to Eq. (5), we obtain for $y_1$ and $y_2$

$$y_1=\frac{a_1}{\omega_1 \sin\varphi_0}\,\frac{\sin\omega_1(\varphi_0-\varphi)}{\sin\varphi},$$

and

$$y_2=a_2C_2\frac{\sin\omega_2(\varphi'-\varphi_i')}{\sin\varphi'} \quad (12)$$

where $\varphi_i'$ is chosen in accordance with Newton's image equations, which hold for this type lens, and where $C_2$ is to satisfy the matching condition.

Slightly modifying Glaser's results we then obtain

$$\frac{\Delta r}{M}=\alpha^3\left[\frac{a_1k_1^2}{6\omega_1^4}\,\frac{1}{\sin^4\varphi_0}K_1+\frac{a_2k_2^2}{6}C_2^4K_2\right]$$
$$=\alpha^3C_{sp}. \quad (13)$$

For $K_1$ and $K_2$ we write (see Eq. (8))

$$K_1=\int_{\varphi_0}^{\varphi_3} f(k_1,\varphi,\varphi_0)d\varphi,$$
$$K_2=-\int_{\varphi_i'}^{\varphi_3'} f(k_2,\varphi',\varphi_i')d\varphi'. \quad (14)$$

As we have already seen, (Eq. (9)), the exact expressions for $K_1$ and $K_2$ are exceedingly cumbersome; but fortunately they may be approximated—for the values we use—to within a few percent by the following equations:

$$K_1=\tfrac{3}{2}\omega_1^2(\varphi_3-\varphi_0)+\frac{3\,\omega_1^4}{4\,k_1^2}\,\frac{4k_1^2-3}{4k_1^2+3}\sin 2\varphi_0,$$

$$K_2=-\left[\tfrac{3}{2}\omega_2^2(\varphi_3'-\varphi_i')+\frac{4k_2^2-3}{4k_2^2}\omega_2\sin 2\omega_2(\varphi_3'-\varphi_i')\cos 2\varphi_3'-\frac{4k_2^2-3}{4k_2^2}\cos 2\omega_2(\varphi_3'-\varphi_i')\sin 2\varphi_3'\right]. \tag{15}$$

Now we but need an evaluation of $C_2$ to have an analytic expression for the aberration constant, $C_{sp}$. To do this it is convenient to write $y_2$ as

$$y_2=a_2C_2\frac{\sin\omega_2(\varphi'-\varphi_0')}{\sin\varphi'}, \tag{16}$$

$\varphi_0'$ being the position of the virtual image formed by the first lens as seen by the second, and again being determined by the Newtonian image relations.

Although we must match both $y_1$ and $y_2$ and their slopes, it is obvious, $\varphi_0'$ having been determined, that if $C_2$ satisfies either condition it will automatically satisfy the other. We thus pick as most convenient the condition that the path distances (see Eq. (12)) be equal. We have then:

$$a_2C_2\frac{\sin\omega_2(\varphi_3'-\varphi_0')}{\sin\varphi_3'}=\frac{a_1}{\omega_1\sin\varphi_0}\,\frac{\sin\omega_1(\varphi_0-\varphi_3)}{\sin\varphi_3}. \tag{17}$$

Fortunately, the arguments are such that we can approximate as follows:

$$a_2C_2\frac{\omega_2(\varphi_3'-\varphi_0')}{\pi-\varphi_3'}=\frac{a_1}{\omega_1\sin\varphi_0}\,\frac{\pi-\omega_1(\varphi_0-\varphi_3)}{\varphi_3}. \tag{18}$$

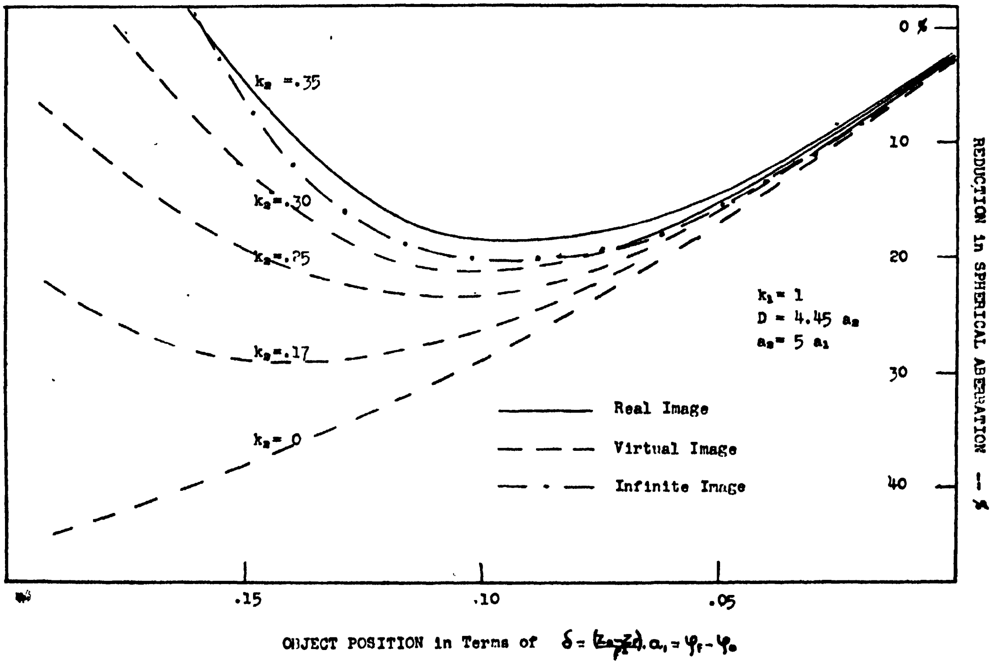


FIG. 4. Spherical aberration *vs.* objective position in first lens, for a second lens of varying strength, but fixed half-width. An infinite image is formed when the first lens forms a virtual image on the focal position of the second lens. Above this boundary, the system forms a real final image, below, a virtual one.

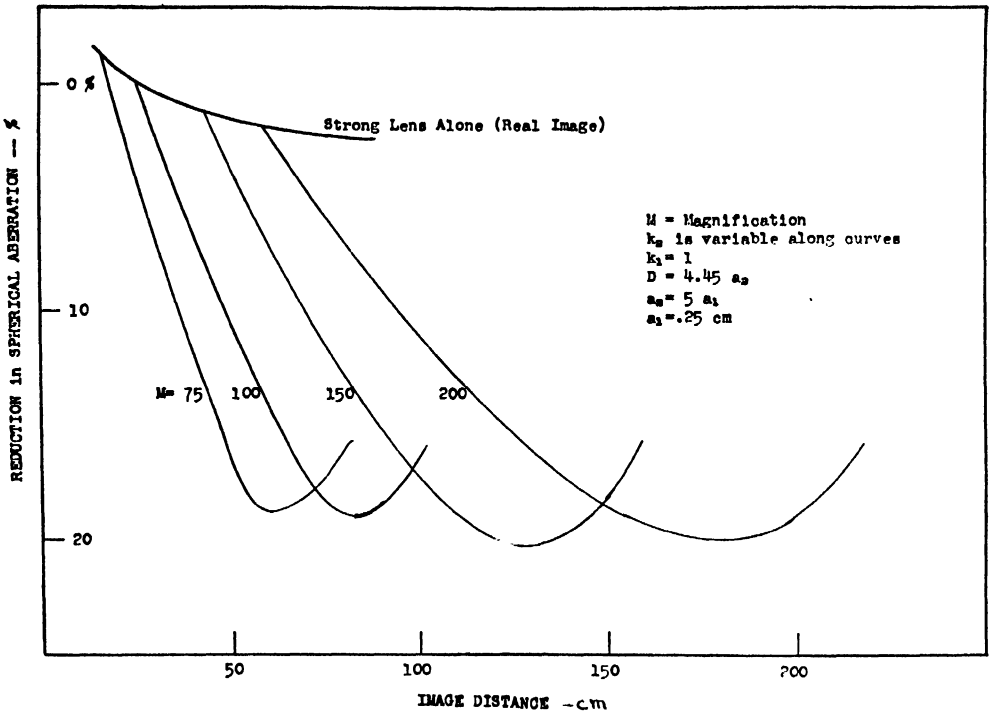


FIG. 5. Spherical aberration *vs.* the final image distance for a typical set of conditions.

Now setting $\varphi_0=\varphi_F-\delta$, and applying the fact that $\omega\varphi_F=\pi$ (Eq. (6)), $\varphi_F$ being the focal position, we obtain:

$$C_2=\frac{a_1}{a_2}\frac{1}{\sin\varphi_0}\frac{1+\delta/\varphi_3}{\omega_2}\frac{\pi-\varphi_3'}{\varphi_3'-\varphi_0'}. \tag{19}$$

Thus we have obtained for the aberration constant of a two-component compound lens:

$$\frac{C_{sp}}{a_1}=\frac{1}{6\sin^4\varphi_0}\left[\frac{k_1^2}{\omega_1^4}K_1+\left(\frac{a_1}{a_2}\right)^3K_2\frac{k_2^2}{\omega_2^4}(1+\delta/\varphi_3)^4\left(\frac{\pi-\varphi_3'}{\varphi_0'-\varphi_3'}\right)^4\right]. \tag{20}$$

## DISCUSSION

We have already seen that $\varphi_0$ should be close to $\pi/2$. The question remains how the second term varies with $\varphi_0$. As the beam leaving the first lens is converged less, or $\varphi_0\rightarrow\pi/2$, one can expect it to increase rapidly, since the aberration increases with the cube of the aperture. The behavior with $\varphi_0$ for different $k_2$ is shown in Fig. 4.

For the same reason the two lenses should be close together. It is unfortunate that we have not succeeded so far in treating the case of interacting fields, so that we could find the exact behavior when $D$ becomes small. It is certain, however, that practical compound lenses will need to be close enough together to make field interaction inevitable.

One reason for the need is the final image distance involved. Figure 5 gives an idea of the increase in image distance for only a 20 percent reduction in the aberration. Decreasing $D$ is the only way of decreasing the aberration and keeping a reasonable image distance.

Another way of decreasing $C_{sp}$ is by increasing the ratio of the half-width of the second lens to that of the first. (See Fig. 6.) But here again we increase the final image distance, as is also indicated on the graph, in a manner approximately proportional to $a_2/a_1$.

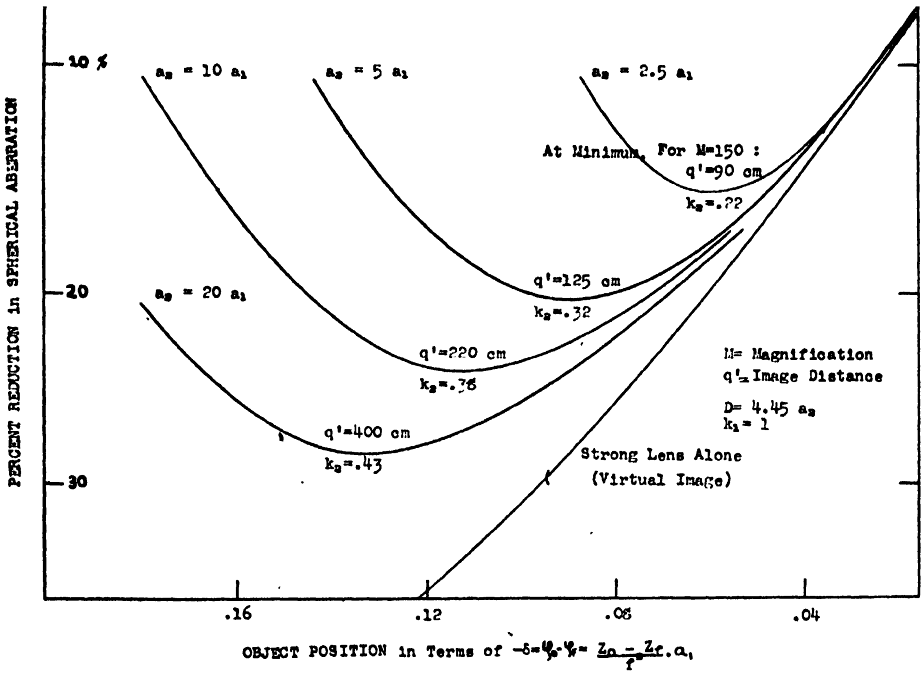


FIG. 6. Spherical aberration *vs.* object position, for a second lens of varying half-width. The total magnification along the curves is infinite, but to gain an idea as to how the image distance $q'$ varied, we took the lens strength indicated by the aberration minimum and calculated $q'$ for a magnification of 150.

Thirdly, the contribution of the second lens is almost proportional to $k_2^2 (k_2^2 \ll 1)$, shown in Fig. 4. There again, however, another consideration enters. The focal length of the second lens is inversely proportional to $k_2^2$. The greater this focal length, the higher must be the magnification by the first lens in order to produce a virtual image outside the second focal point and thus secure a real final image.

But for high magnification by the first lens one achieves almost no reduction in the aberration. The result is that there is an optimum value for $k_2$, as is shown in Fig. 7 (which is merely Fig. 4 plotted against total magnification).

This optimum value, we note, is not necessarily the best from the point of view of image distance. Referring once more to Fig. 5, it is seen that especially for high magnification the aberration minimum is very flat. One can thus gain a considerably shortened image distance, without appreciably increasing the aberration, by taking a value of $k_2^2$ somewhat less than the optimum value.

We saw from Fig. 4 that some decrease in the aberration is possible using only two lenses, but that even greater decreases come with the formation of a virtual image by a very weak second lens. The question then arises whether one cannot use three lenses to good advantage.

Our next problem is to decide therefore whether the contribution of the third lens can be made sufficiently small to make its addition worth while.

*Case III. Three lenses.* Again proceeding as in the previous section we easily obtain for the aberration:

$$\genfrac{}{}{0pt}{}{C_{sp}}{a_1} \quad \frac{\Upsilon}{6\sin^4\varphi_0}\left\{\frac{k_1^2}{\omega_1^4}K_1+\left[\left(\frac{a_1}{a_2}\right)^3\left(1+\frac{\delta}{\varphi_3}\right)^4 K_2\frac{k_2^2}{\omega_2^4}\left(\frac{\pi-\varphi_3'}{\varphi_3'-\varphi_0'}\right)^4\right]\right.$$

$$\left.\times\left[1+\left(\frac{a_2}{a_3}\right)^3\frac{K_3}{K_2}\cdot\frac{k_3^2}{k_2^2}\frac{\sin^4\omega_2(\varphi_4'-\varphi_0')}{\sin^4\omega_3(\varphi_4''-\varphi_i'')}\frac{\sin^4\varphi_4''}{\sin^4\varphi_4'}\right]\right\}. \quad (21)$$

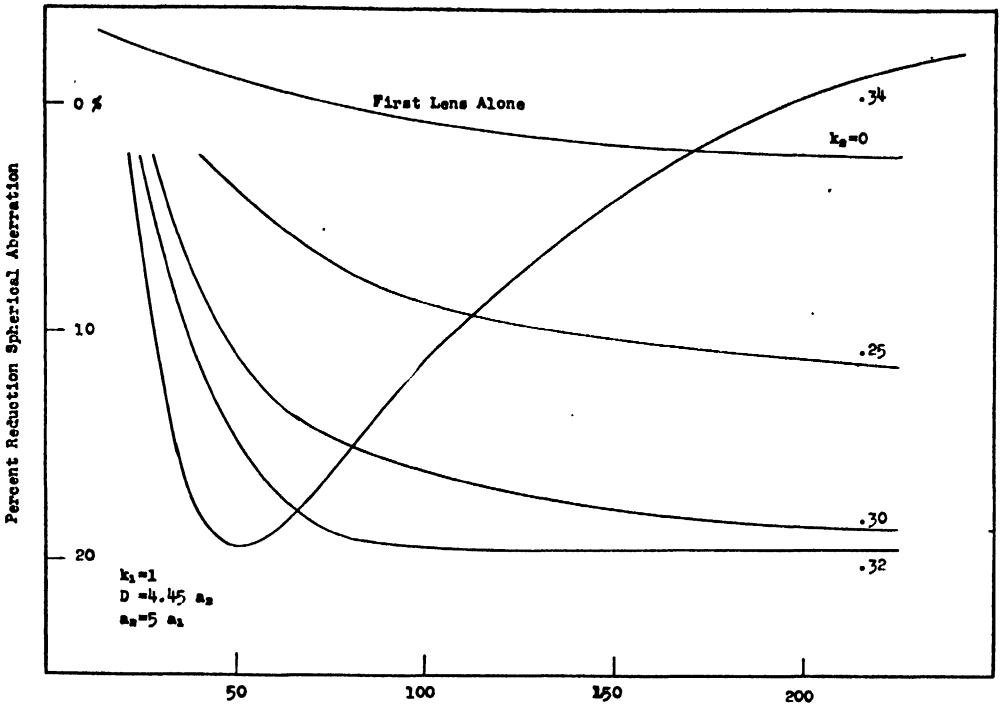


FIG. 7. Aberration *vs.* total magnification. This is a plot of graph II *vs.* magnification instead of object position.

The symbolism is exactly analogous to that for Case II except that the subscript 4 refers to the field-free plane between the second and third lenses. All double primes refer to the coordinate system of the third lens.

Equation (21) leads us to the following qualitative conclusions:

(1) By making $a_2/a_3$ small enough (say $\frac{1}{5}$) we can obtain $C_{sp_{3\text{ lenses real image}}} = C_{sp_{2\text{ lenses virtual image}}}$. In other words, the last term in Eq. (21) must be $\ll 1$.

(2) The image distance, $q$, depends almost entirely on the spacing between the last two lenses. If we wish to satisfy the condition of no interaction between the lenses, $q$ will be greater than for the two lens combination.

Both aberration and image distance, for the combination of two or three lenses, show a considerable improvement only for conditions exceeding the limits imposed by our method of derivation. It is hoped that a more rigorous derivation and/or experiments will enable us in the near future to substantiate and to extend the results of this preliminary calculation.

## CONCLUSION

By combining two lenses one can approach the lower limit to the aberration set by the first lens, used as a virtual image former. For such lens separation for which the interaction of the lens fields is negligible, the image distance is considerably greater than that of the first lens used as a real image former.

How closely one can approach this lower limit cannot be answered without a study of the aberration for very close, interacting fields.

Certainly it seems desirable to have the lenses close together.

In addition it has been found that there is an optimum value for the strength of the second lens in a two-lens system, and that the half-width of the second field should be several times greater than that of the first.

Three lenses may be used to improve the aberration, but only at a cost of a great image distance.

# Physical Properties of Calcium Tungstate X-Ray Screens

J. W. COLTMAN, E. G. EBBIGHAUSEN,* AND W. ALTAR
*Westinghouse Research Laboratories, Pittsburgh, Pennsylvania*

(Received January 6, 1947)

The absolute energy of the light from a fluorescent calcium tungstate x-ray screen, its spectral distribution, and the efficiency of conversion from x-rays to light have been determined. The absolute energy and its spectral distribution were obtained spectrographically by comparison with a standard tungsten filament lamp. Equations are given for determining the distribution of light in the diffusing screen material, and calculation of the light losses are made for varying screen thickness and x-ray absorption. The data for energy distribution of the ingoing x-rays were taken from measurements obtained in this laboratory by a new method to be published elsewhere. Allowance for scattering and secondary *x*-radiation was made and a conversion efficiency of 5.0 percent was obtained for this calcium tungstate phosphor.

## INTRODUCTION

THE primary purpose of this research was to obtain experimentally the absolute efficiency of the transformation of x-ray energy into light energy by calcium tungstate. Of the various fluorescent materials which might have been examined, calcium tungstate appeared to be the one whose investigation would be most fruitful, first, because of its wide utilization in radiography, and second, because the fluorescence of this phosphor is not due to the presence of a foreign material serving as an activator. Thus calcium tungstate phosphors produced by different manufacturers should have the same spectral distribution, and probably will not vary greatly in their efficiency. As a basis for theoretical studies of fluorescence, knowledge of the behavior of calcium tungstate is probably of more value than that for the sulphide phosphors which are very sensitive to minute quantities of impurities.

The problem of determining the conversion efficiency divides itself into three main parts. First, the absolute energy and spectral distribution of the light emitted from the screen must be determined. Second, the light losses in the screen must be determined in order that the total light generated may be calculated. Third, the energy of x-rays absorbed in the screen must be measured. The resultant efficiency will be a characteristic of the calcium tungstate itself and of the wave-length of the exciting x-rays. With this information and a knowledge of the functional dependence of the light losses on the structure of the screen as presented herein, the behavior of any calcium tungstate screen may be predicted. The method of analysis developed to obtain the light losses are applicable to the transmission of light in diffuse media in general and, in particular, to phosphors other than calcium tungstate.

## Ia. LIGHT ENERGY AND ITS SPECTRAL DISTRIBUTION

The spectrographic part of the problem consisted of the determination of the spectral energy distribution of the $CaWO_4$ screen when excited by 80 kv x-rays, filtered by one half millimeter of Cu and one millimeter of Al, and with a tube current of 20 milliamperes. The procedure involved a photographic comparison between the known spectral energy distribution of a temperature-calibrated tungsten ribbon filament lamp and that of the $CaWO_4$ screen.

### The Experimental Apparatus

The standard source was a lamp with a tungsten ribbon filament and quartz window loaned by the National Bureau of Standards. The ribbon filament is about $1\frac{1}{2}$ inches long and $\frac{1}{8}$ inch wide and is supported in the bulb by heavy nickel leads.

The temperature calibration of the lamp was made by Mr. M. S. Van Dusen, Acting Chief of the Pyrometry Section, who gave the value of the brightness temperature outside of the quartz

---

* Now at University of Oregon, Eugene, Oregon.

window as 2614°K at a current of 39.20 amperes and for a point about midway along the filament. The position is marked by a small nick on one edge of the filament itself. Values of the spectral transmission of the quartz window were also supplied by Mr. Van Dusen.

In order to compute the spectral distribution of the radiation from the filament, it was necessary to obtain the true temperature of the filament inside the bulb and to have a knowledge of the spectral emissivity of tungsten over the intended working range of wave-length (3000A–7000A). A compilation of spectral emissivities was furnished by Dr. W. W. Coblentz, retired head of the Radiation Laboratory of the National Bureau of Standards. The conversion from the brightness temperature to the true temperature is accomplished by the following formula, derived from Planck's radiation formula:

$$\frac{1}{T_t}=\frac{1}{T_{BO}}+\frac{\lambda}{.4343C_2}\log_{10}(t_\lambda\cdot\epsilon_\lambda), \qquad (1.1)$$

where

$T_t$ = true temperature of the filament

$T_{BO}$ = brightness temperature of the filament outside the window as determined by the N.B.S. optical pyrometer

$\lambda$ = mean effective wave-length in cm of the filter used in the N.B.S. optical pyrometer

$C_2$ = the second constant in Planck's radiation formula

$t_\lambda$ = transmission of the quartz window at the above-mentioned wave-length

$\epsilon_\lambda$ = the spectral emissivity of tungsten at $\lambda$.

With $T_{BO}=2614°\text{K}$, $\lambda=6650\times10^{-8}$ cm, $C_2=1.435$, $t_\lambda=0.927$, $\epsilon_\lambda=0.426$, the true temperature of the filament is 2947°K.

The spectral energy distribution outside the lamp was obtained by substitution of the above value of the true temperature in Planck's radiation formula, and correction of the result for the spectral emissivities of tungsten and for the spectral transmission of the quartz window.

When the image of the central portion of the filament was focussed on the slit of the spectrograph, the resulting spectrum was a continuous band with a narrower and darker band running down its length. This latter band was caused by the region of the filament at the nick where the filament was narrower, and its resistance and hence its temperature higher, than regions of the filament above and below the nick. This narrow band was not wide enough to measure in the microphotometer, so it was necessary to use some other portion of the filament where the temperature variation with length was small. Measurements of temperature with an optical pyrometer were made covering the $\frac{1}{2}''$ range from a point $\frac{1}{8}$ inch above the nick to within $\frac{1}{8}$ inch of the top of the filament. The true temperature fell only about seven degrees over this range, and at its center was 41° lower than that at the nick. The portion of this region used for projection onto the length of the slit (2 mm) of the spectrograph was about $\frac{3}{8}$ inch. Hence the true temperature variation over this range was of the order of about 4°. From Planck's radiation law one derives the percentage change in energy at a given wave-length for a change in true temperature to be

$$(dE_\lambda/E_\lambda)\ \text{percent}=100(C_2/\lambda T^2)dT. \qquad (1.2)$$

For a true temperature of 2906°K the percentage change of $E_\lambda$ per degree is about 0.57 percent at 3000A and 0.28 percent at 6000A. One may take 0.4 percent as a mean. Hence, a change of 4° (true) corresponds to less than a 2 percent change in $E_\lambda$. Furthermore, since the quartz lens used to focus the filament on the slit of the spectrograph does not give a perfectly sharp image, the effect is to smooth out variations in $E_\lambda$ and produce a spectrum of uniform density across its width. Instead of recomputing the energy distribution for the adopted temperature of 2906°K, the one for 2947°K was used with the correction factor

$$\begin{aligned}C&=\exp(C_2/\lambda)(T_1-T_2/T_1T_2)\\&=\exp(C_2/\lambda)(\Delta T/T^2), \qquad (1.3)\end{aligned}$$

where $\Delta T=41°$ and $T$ is the mean of 2947°K and 2906°K. It appears that the combined error from all sources in the computed energy varies from about 6 percent at 3000A to about 3.0 percent at 6500A with an average for the middle of the spectrum of about 4.5 percent.

One of the most difficult problems was the reduction of the filament brightness to a value closely approximating that of the fluorescent screen. This was necessary because of reciprocity failure of the photographic emulsion which demands that the exposure times on the standard lamp and the fluorescent screen be of quite the

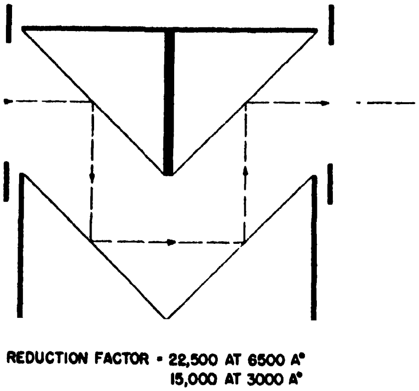


FIG. 1. Schematic arrangement of reflection-type attenuator.

same order of magnitude. A rough computation on the basis of available data indicated that the lamp might be brighter by a factor of the order of one million. Only by the combined use of three methods could the filament brightness be reduced to the point where the exposure times became nearly equal. Reduction by transmission filters had to be rejected as impractical, because transmission coefficients were not known with sufficient accuracy at the various wave-lengths. Of the following three methods used, one will be discussed in detail now and the second and third in later sections:

1. The method of successive reflections.
2. Rotating sectors.
3. Stops in the collimator of the spectrograph.

In the first method the brightness was reduced by successive reflections from four air-glass surfaces. The apparatus designed and constructed for this purpose was called the "attenuator." The attenuator consisted of four borosilicate crown glass prisms arranged as in Fig. 1. The arrows indicate the path of the light through the system. The system was enclosed in a brass box, and the prisms were carefully aligned so that parallelism and perpendicularity of the faces was assured. The angle of incidence of the central ray of the beam closely approximated 45° at all of the four surfaces of reflection. At each surface a small fraction of the light was reflected, and the remainder was transmitted into the prism. In order that as little as possible of this light should be returned to the beam by internal reflections, all prism surfaces except the ones used for reflection were coated with a solution of propyl methacrylate mixed with lampblack, and the prisms were baked in an oven at 120°F for two hours. The result was a non-reflecting and absorbing coating which served to remove to a very high degree all light transmitted into the prism.

According to Fresnel's formulas the amplitude ratios of the reflected light are:

$$\frac{\sin(I-I')}{\sin(I+I')} \quad \text{and} \quad \frac{\tan(I-I')}{\tan(I+I')}$$

for light polarized in, and, respectively, at right angles to, the plane of incidence. The intensity of the parallel component after four successive reflections becomes:

$$R(I_{\parallel})=\left[.707\left(\frac{\sin(I-I')}{\sin(I+I')}\right)^4\right]^2 =\frac{1}{2}\left(\frac{\sin(I-I')}{\sin(I+I')}\right)^8, \qquad (1.4)$$

and similarly for the perpendicular component:

$$R(I_{\perp})=\frac{1}{2}\left(\frac{\tan(I-I')}{\tan(I+I')}\right)^8. \qquad (1.5)$$

Computation shows that for the range of spectrum to be studied, $I_{\perp}$ is about 9000 times fainter than $I_{\parallel}$ and hence the light may be said to be completely polarized.

With $I$ fixed at 45°, $I'$ depends upon $n$ (Snell's Law) and since $n$ varies with the wave-length, $R$ itself is a function of the wave-length.

In order to determine the variation of $n$ with wave-length, one of the prisms was mounted on a spectrometer and $n$ was determined for each of the three mercury lines of wave-length 5460.7A, 4358.3A, and 4046.6A. The resulting indices were, respectively, 1.5260, 1.5349, and 1.5389. Each determination had a standard deviation of unity in the fourth place of $n-1$. With these three values of the wave-length and their respective indices, the constants of the Hartmann formula were obtained:

$$n=1.5041+\frac{83.416}{\lambda-1648.0}. \qquad (1.6)$$

For several wave-lengths between 3000A and 6500A, $n$ was determined from (1.6), then $I'$ from Snell's Law and finally the reduction factors from (1.4). The curve of reduction factor *vs.* wave-length is given in Fig. 2. These factors were applied to the spectral energy distribution curve of the filament outside the lamp and result in the distribution curve of the light as it leaves the attenuator.

Attention must be given to the fact that the light from the attenuator is practically completely polarized, while the light from the fluorescent screen is unpolarized. Thus the loss in the spectrograph may not be the same for both sources. In order to test this, two spectrograms of equal exposure times were taken of the lamp through the attenuator. For the first exposure, the plane of the normals to the reflecting surfaces was held perpendicular to the length of the slit and parallel for the second exposure. From the resulting densities of the spectra it appeared that the effective exposure was only about half as great in the first case as in the second. In order to avoid the difficulty the attenuator was mounted so that the plane of the normals made an angle of 45° with the slit and, therefore, as far as the spectrograph was concerned, it was receiving isotropic light.

The errors in the computed reduction factors are evaluated as follows:

1. As already stated $n$ has been determined with a standard deviation of one part in the fourth place of $n-1$. For the middle of the spectral range, this value results in an error of about 0.1 percent in the value of $R$— a negligible quantity.
2. Since $n$ was determined for only three wave-lengths, interpolation had to be used for shorter and longer wave-lengths using formula (1.6). The error in $R$ to be expected, particularly in the short wave-length region where $n$ is changing rapidly, is estimated to be of the order of 0.3 percent in the neighborhood of 3200A and of 0.1 percent for wave-lengths longer than 4000A.
3. The beam through the attenuator, as defined by the lens which focusses the filament on the slit and by the angular size of the collimator, has an angular diameter of about two degrees, of which only the central ray will strike the reflecting surfaces at an angle of incidence of 45°. The limiting values at the first surface will be 44° and 46°. However, it will be seen that (by design) if $I$ is 46° at the first surface, it will have the same value at the second, 44° at the third, and 44° at the last surface. Calculation shows that the error made in $R$ by assuming $I=45°$ for the whole beam is about 0.2 percent and thus is negligible.

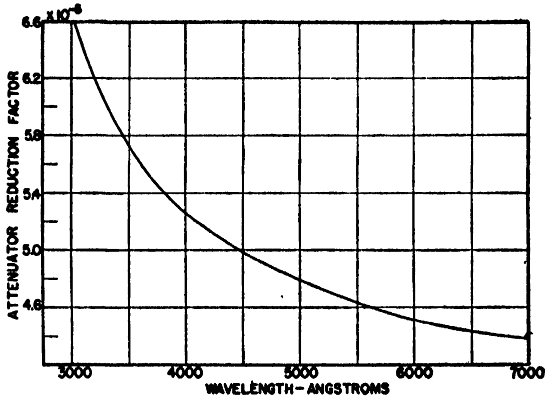


FIG. 2. Reduction factor of attenuator as a function of wave-length.

4. Particular care was used in the alignment of the attenuator so that the resulting error in $R$ is probably less than 0.4 percent.

The total error is probably not in excess of 1 percent.

The spectrograph was an all quartz optics, low dispersion instrument. The arrangement was such that the plate holder could be moved at right angles to the spectrum, and thus several spectra could be obtained on one plate. The length of spectrum from 3000A to 6500A was about $\frac{3}{4}$ inch. The spectrograph was mounted on a turntable which permitted the slit to be turned to either the standard lamp, the mercury lamp (for a comparison spectrum), or the fluorescent screen. By means of lead sheets, the whole instrument was very well shielded from the direct or scattered radiation of the x-ray tube. On the turntable was mounted a motor to turn cardboard sectors, used to reduce the lamp brightness by a given amount, and also a mount for the quartz lens to focus the lamp filament on the slit. Inside the prism case was an arrangement to hold stops for the further control of the lamp brightness. The length of the slit was 2 mm and the width 0.2 mm.

The fluorescent material, a sample of Eastman Ultra Speed back screen, consisted of a layer of $CaWO_4$ covered by a film of lacquer and mounted on a sheet of white cardboard four inches square. It was centered over a two inch square hole in a sheet of $\frac{1}{8}$ inch lead and the whole unit was mounted over the open end of the protective cone of the x-ray tube.

## The Experimental Procedure

Generally speaking, our procedure was a comparison, at a number of wave-lengths, between the brightness of the screen and the standard lamp filament as seen through the attenuator. Given the magnitude of the latter and its spectral distribution, it was possible to compute the light output of the screen on an absolute scale.

One of the most difficult problems, the reduction of the brightness of the standard lamp to a level comparable to that of the fluorescent screen, was accomplished by the use of the attenuator (already described), rotating sectors, and collimator stops. This permitted the use of equal exposure times for the lamp and the fluorescent screen and thereby avoided the danger of reciprocity failure. In order to provide a calibration on each plate, five exposures were made using five cardboard sector disks, eight inches in diameter; the first, *A*, providing a reduction factor 12.5 and the others, *B*, *C*, *D*, and *E* approximately 25, 50, 100, and 200. When using these disks, one must avoid the intermittency effect because of which the photographic density is a measure of the total light impinging only under certain conditions. A continuous exposure and an intermittent one of the same total energy will produce the same density on the photographic plate, provided the total time of exposure is the same in both cases, and the frequency of interruption is above a certain critical rate. The Eastman Kodak Company's publication, "Photographic Plates for use in Spectroscopy and Astronomy," states that this critical rate varies from about 100 flashes per second for a one-second exposure to a few flashes per second for a one hour exposure. The motor used to drive the sectors operated at 2850 r.p.m. and since no exposure was less than twenty minutes, no error caused by the intermittency effect was incurred.

The ~~final ad[illegible]ment of~~ lamp brightness was obtained by placing one of several stops between the prism and the collimator lens at a position close to the latter. The first stop defined the beam leaving the collimator lens to the extent that all of the beam entered the prism and passed through the camera lens to the photographic plate. This stop was used when photographing the spectrum of the fluorescent screen, and its reduction factor was designated as unity. Six more stops were made with reduction factors varying from 2 to 37. With this last stop, the total reduction of the lamp brightness, including the average of that of the attenuator and that of sector *A*, reached a factor of nine million. This much was needed when working in the region around 6000A where the screen fluorescence is weak, and the lamp output high.

All photographic plates used were coated with Eastman 103-F emulsion. On each plate eight exposures were made; one from the fluorescent screen, five from the standard lamp using sector disks *A* to *E*, and two exposures from a mercury lamp to obtain wave-length scales at the top and bottom of a plate.

The densities of the spectra were compared in an Arl Dietert microphotometer. At each position the density was determined for the fluorescent spectrum and for each of the five comparison spectra.

Since different portions of the spectrum required widely varying exposure times to give densities on the linear portion of the emulsions, it was necessary to use several plates to cover the complete spectrum. For each wave-length the transparencies of the five comparison spectra were plotted on semi-logarithmic paper against the reduction factors of the sectors, taking that for sector *A* as unity. From this curve was determined the factor by which the fluorescent spectrum was fainter than the spectrum belonging to sector *A*.

This factor was combined with reduction factors for the collimator stop, the factor 12.5 for the sector disk *A* and the factor for the attenuator, to give a total reduction ratio between intensities at corresponding wave-lengths from the standard lamp and from the $CaWO_4$ screen. A total of seven plates were measured.

## Experimental Results

Near the end of the investigation it was found that the portion of the $CaWO_4$ screen which had been exposed to x-rays was somewhat fainter, showing that it had suffered a certain degree of "fatigue." Measurements indicated that the exposed portion of the screen was about 8 percent fainter than the outer portion that had been protected from x-radiation by the lead.

In order to deduce from the measurements on a fatigued screen the spectrum of a fresh screen, measurements on each plate were corrected by an amount proportional to the total x-ray exposure time of the screen up to the time when the plate in question was taken.

The total time of exposure of the screen to x-rays was about six and one-half hours. This includes exposure times for test plates and several others that for one reason or another were not used. The results from the individual plates are given collectively in Fig. 3. The ordinates are $10^{-5}$ ergs/cm$^2$ sec. *A* unit. The assumption is made that both the lamp filament and fluorescent screen radiate according to Lambert's cosine law, and a correction for departures will be added later.

The curve given in Fig. 3 is not final, but must be corrected for slit broadening. The effect is to pull in both ends of the curve to a certain extent without changing the total area. Integration of this curve yields the result that the total energy output of the screen was 2.54 ergs/cm$^2$/sec.

A correction factor of 0.92 must be applied to this value to correct for the departure from Lambert's law in the angular distribution of the light from the screen. The experimental determination of this factor is described in Section Ib. The total energy thus reduces to 2.34 ergs/square centimeter per second, for a measured x-ray intensity of 0.149 *r*-units per second incident on the screen. The final corrected spectral curve, Fig. 4, is given in terms of light energy generated in the screen per *r*-unit absorbed. The method for measuring the light losses is given in Section 2.

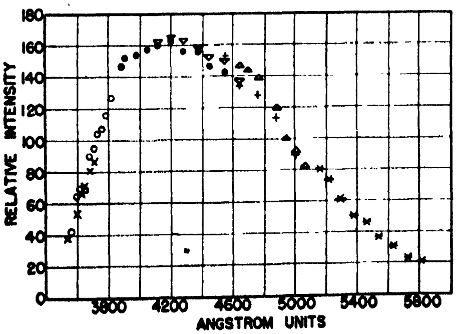


FIG. 3. Spectrum of CaWO$_4$ as obtained from seven plates (uncorrected).

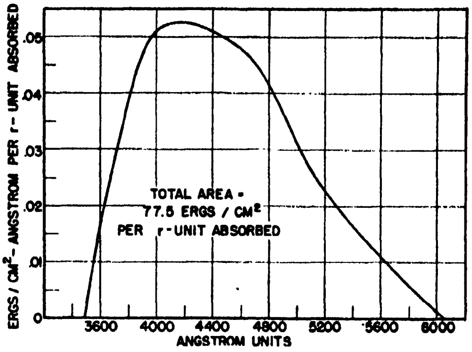


FIG. 4. Fluorescent spectrum of CaWO$_4$ excited by 80 Kv constant potential x-rays. Filter used was ½ mm copper+1 mm aluminum. Screen contained .109 gm/cm$^2$ of CaWO$_4$.

An estimate of the accuracy of this determination involves a consideration of such sources of error as the temperature of the lamp, the spectral emissivities of tungsten, the brightness reduction factors, and errors of the photographic emulsion. It appears that the final total energy has been determined to within an accuracy of about 12 percent.

### Ib. ANGULAR DISTRIBUTION OF THE LIGHT

In order to calculate the total light emitted from the screen from the measured brightness in a specific direction, it is necessary to know the angular distribution of the emission from the screen. In the case of the filament, this factor was taken care of by using the normal emissivity of tungsten for calculation. For the screen, measurement of the angular distribution was made by means of a photo-multiplier cell illuminated through a collimator consisting of a simple lens and a small diaphragm stop at its focal point. The whole assembly was capable of rotation about an axis lying in the plane of the fluorescent screen.

If the emission of light followed a Lambertian (cosine) distribution, the light received by the photocell would be independent of angle. Actually a falling-off at large angles was observed as shown in Fig. 5. Integration of this curve combined with suitable angular variables yields a constant which relates the total light emitted from the screen to that which would have been emitted from a Lambertian source with the same normal intensity. This factor varies slightly with

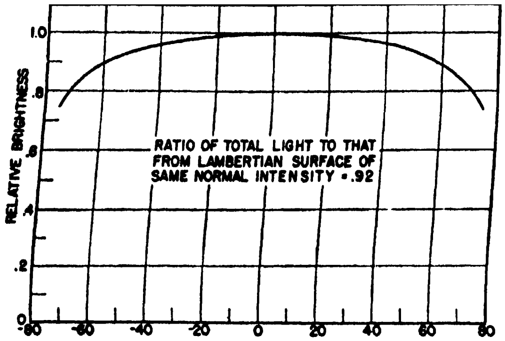


FIG. 5. Angular distribution of light from $CaWO_4$ screen relative to a Lambertian surface.

the wave-length of the exciting x-rays, since the latter affects the distribution of light-generating centers within the screen. Under the conditions of the spectroscopic experiment the $CaWO_4$ screen had a factor of 0.920.

## II. TRANSMISSION OF LIGHT IN DIFFUSE SCREENS

The light produced by the x-rays in the interior of the fluorescent layer is eventually transmitted out of the screen by means of a series of reflections and refractions from individual crystals. Part of the light is lost by absorption in the crystals and in the binder, and part finds its way out of the back surface of the layer. Some of the latter may be recovered by means of a reflecting backing, a white mounting card being commonly used.

Since the particles involved are small compared to the area of interest, and the lateral dimensions of the screen are large compared to the thickness, the problem can be treated as if the material were homogeneous and as if the light were traveling in one dimension only, i.e., perpendicular to the screen.

In the following discussion the words "light forward" and "light backward" refer to directions in the same sense as that of the x-rays. Thus with the screen located between the x-ray tube and the observer, the light forward is that seen by the observer and comes from what will be termed the front of the screen; the light backward from the back of the screen is directed toward the x-ray tube.

The first step in the analysis is to derive the reflection and transmission coefficients of a diffuse layer as a function of its thickness. A layer of thickness $x$ (Fig. 6) has by definition a reflection coefficient $R(x)$. If a layer of thickness $dx$ is added to this, a new reflection coefficient will result. The reflection coefficient of the elementary layer itself will to the first order be proportional to its thickness. Let this reflection coefficient be $kdx$. Similarly, the amount *not* transmitted will be proportional to the thickness to the first order; let this coefficient be $\alpha dx$. It will be evident that $\alpha \geq k$, being equal to $k$ if there is no true absorption of light in the layer. (In Fig. 6 the layer $dx$ is shown separated from the layer $x$ for convenience in tracing the course of the light. It is to be understood that in this and later derivations the layers are actually continuous.). Light is incident on the combination from the left. A fraction $kdx$ will be reflected from the first layer, a portion $1-\alpha dx$ will be transmitted to the second layer. The amount reflected from the second layer is $(1-\alpha dx)R$, of which a quantity $(1-\alpha dx)^2R$ is transmitted out to the right and adds to the initially reflected light. The portion $kdx(1-\alpha dx)R$ which is returned undergoes an exactly similar set of steps. Summing up the infinite series of rapidly diminishing contributions:

$$kdx+R(1-\alpha dx)^2(1+Rkdx+R^2(kdx)^2+\cdots)$$

$$=kdx+\frac{R(1-\alpha dx)^2}{1-Rkdx}=R(x+dx), \quad (2.1)$$

the new reflection coefficient $R(x+dx)$ is obtained. Subtracting $R$ from both sides, and

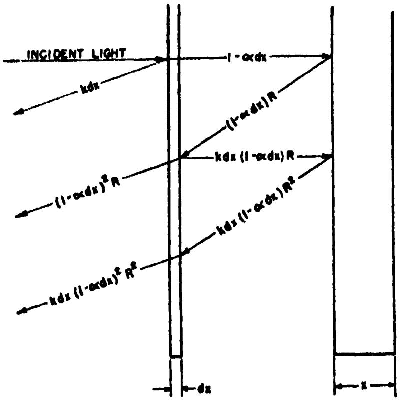


FIG. 6. Schematic representation of light flow for derivation of reflection coefficient formula.

neglecting differentials of higher order than one, we obtain

$$R(x+dx)-R(x)=k(1-2(\alpha/k)R+R^2)dx$$

or

$$\frac{dR}{dx}=k(1-2(\alpha/k)R+R^2).$$

Thus:

$$kx=\int\frac{dR}{[1-2(\alpha/k)R+R^2]},$$

$$kx=-\frac{1}{[(\alpha/k)^2-1]^{1/2}}\tanh^{-1}\frac{R-\alpha/k}{[(\alpha/k)^2-1]^{1/2}}+\text{const.},\tag{2.2}$$

$$R=\frac{\alpha}{k}+[(\alpha/k)^2-1]^{1/2}\tanh((\alpha^2-k^2)^{1/2}x+\text{const.}).$$

The constant is determined by setting $R=0$ when $x=0$, and by further reduction we obtain

$$R=\frac{(k/p)\sinh px}{\cosh px+(\alpha/p)\sinh px},\tag{2.3}$$

where

$$p\equiv(\alpha^2-k^2)^{1/2}.$$

By similar consideration of the multiple reflections in the transmission case, and by use of formula 2.3, the expression for the transmission coefficient is found to be

$$T=\frac{1}{\cosh px+(\alpha/p)\sinh px}\tag{2.4}$$

where $p$ has the same meaning as before.

The constant $p$ is an important characteristic of the material and represents the degree of absorption. For very large $x$, 2.4 reduces to

$$T=\text{constant }e^{-px}$$

the usual exponential function, and $p$ plays the role of a linear absorption coefficient. At large $x$, $R$ approaches a constant, and only after this condition is reached will $T$ be exponential. The physical reason for this is clear; if a thin layer is placed behind another layer, the *incident* light on the new layer is enhanced by the multiple reflections between the two, and only if the original layer is so thick as to have attained the limiting reflection coefficient will the transmission be exponential. The form of these functions is shown in Fig. 7 for values of $k$ and $\alpha$ typical of $CaWO_4$. The values of $k$ and $p$ (or $k$ and $\alpha$) suffice to describe completely the light yield from the front and back of the screen, as well as the light absorbed, if the thickness and distribution of light sources within the screen are given. These coefficients may be experimentally determined by measuring $R$ and $T$.

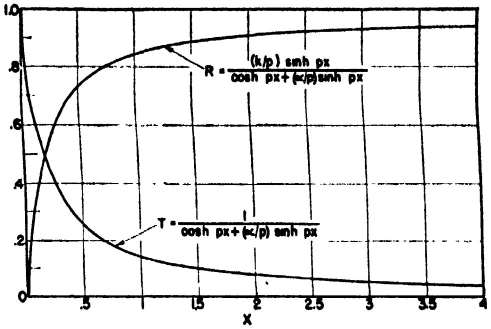


FIG. 7. Reflection and transmission coefficients as a function of thickness ($x$) of diffusing material.

### Determination of the Optical Constants

Values for $p$, $k$ and $\alpha$ could be obtained in principle by measuring $T$ or $R$ as a function of thickness, by virtue of Eqs. (2.3) and (2.4). In carrying out such an experiment one must make sure, however, that the amount of light from the source and the action of the detector for the light are not affected by introduction of the sample. For example, a ground-glass diffuser is not a suitable source because it is itself reflecting; introduction of the sample will enhance the incident light intensity to a value higher than that measured without the sample. For the same reason the detector must be non-reflecting; a photographic plate, for instance, is not sufficiently absorbing to avoid errors due to multiple reflections.

Since the procurement of such ideal light sources and detectors constitutes a serious experimental difficulty, it seemed preferable to eliminate the error altogether by means of a method which does not require measurement of light intensity in the absence of the sample. A tungsten bulb, suitably housed, was used to illuminate one side of the layer; the brightnesses of the two sides were measured by means of a Macbeth Illuminometer and the ratio of the brightnesses was taken to be $R/T$. This was repeated with samples of varying thickness.

By definition, $R$ and $T$ are the results of a scattering process which does not alter the angular distribution of the light. Therefore, Eqs. (2.3) and (2.4) must not be applied indiscriminately to the present case where a nearly parallel light beam is incident on the front of the sample and, upon entering, is gradually scattered in all directions, approaching an equilibrium distribution in direction only as a limiting case. The angular distribution of the light thus varies with depth of penetration, whereas Eqs. (2.3) and (2.4) were derived from an assumed stationary equilibrium for the angular distribution; so the latter is presumed to remain unaffected by further scattering and not to vary with depth.

While this poses a very serious difficulty, generally speaking, for the measurement of arbitrary samples, the calcium tungstate screen has the fortunate quality that its scattering power outweighs its absorbing power by an order of magnitude; consequently determination is possible by examining sufficiently thick samples. If the conversion from parallel light to light with completely random orientation takes place in a relatively thin front layer, further re-distribution will proceed in accordance with the underlying assumptions of Eqs. (2.3) and (2.4).

An empirical criterion of this condition is afforded by inspection of the spatial distribution of the light emerging from the front and back of the screen. For a very thin layer it is clear that while the reflected light has a nearly random distribution, the transmitted light will contain large portions of unscattered light; thus a difference in the angular distribution of the light from the front and the back indicates that equilibrium has not yet been reached. In the thinnest sample used a rough check of the angular distributions showed they were alike and nearly Lambertian. Furthermore, the reflectivity of the thinnest sample was calculated to be 85 percent. Thus addition of further layers can change the reflected light by 10 percent at most and the brightness of the front side is always a good measure of the incident light quantity. For this reason, even though the values of $k$ and $\alpha$ may be somewhat in error due to our use of a parallel beam, this does not apply to the value for $p$, which can be computed from the logarithmic derivative $d\log T/dx$ for large thicknesses $x$.

Samples were prepared by soaking an Eastman Ultra-Speed front screen in alcohol and thus removing its fluorescent layer. This layer contained 0.0445 gm/cm$^2$ of $CaWO_4$ per cm$^2$. It is self-supporting because the internal binder is not soluble in alcohol, whereas the cement used to mount it on the card is soluble. One-inch squares were cut from the layer while still wet, and were superimposed to make up samples of 1, 2, and 4 layers. These were cemented over holes in metal sheets and allowed to dry.

It might be supposed that the protective coating on the layers (a film of cellulose plastic) would violate the assumption of homogeneity of the material. However, consideration of the multiple reflections between the crystalline material boundary and film-air interface show that for the index of refraction of this film the effect is completely negligible for the thicknesses used.

Four points on the $R/T$ curve were determined, at zero, 1, 2, and 4 layers. Values of $R/T$ are shown in Table I. From Eqs. (2.3) and (2.4) we see $R/T$ should be of the form $k/p\,\sinh px$. Values of $k$ and $p$ were chosen to fit the experimental points, namely, $k=5.55$/layer and $p=0.30$/layer. The third column gives the calculated points for comparison. Since $\alpha^2=k^2+p^2$, $\alpha$ is also 5.55/layer within the experimental error, with the understanding of course that

$$(\alpha^2-k^2)^{\frac{1}{2}}=0.30/\text{layer}.$$

From the fit obtained, the values of $k$, $\alpha$, and $p$ are estimated to have probable errors of less than 3 percent. In the following treatment these values have been used, and the thickness $x$ is always measured in terms of one layer as the dimensional unit.

TABLE I.

| $x$ (layers) | $R/T$ (exp.) | 18.5 sinh.30$x$ |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 5.43±.05 | 5.63 |
| 2 | 11.8 ±.13 | 11.8 |
| 4 | 27.6 ±.4 | 27.9 |

After these measurements had been evaluated and the results had been used to compute the conversion efficiency and the curves of light yields, it was decided to repeat the measurements of the optical constants of the screen using a diffuse source for the incident light. The sample was mounted over an aperture in an Ulbricht photometric sphere arranged so that the brightness of both sides of the sample could be observed with a Macbeth Illuminometer. These measurements gave a good fit for the ratio $R/T=k/p\,\sinh px$, using $k=6.0$ and $p=0.33$. Adoption of these new values would change the curves of light yields which are presented below. Since these are meant to be illustrative, and quantitative only pending measurement of the optical constants of an individual sample, it did not seem worth while to recompute them. However, the new constants were used for computation of conversion efficiency.

### The Light Distribution in the Screen

Having derived expressions for transmission and reflection of a diffuse layer, one is in a position to compute, for a given distribution of light generation throughout the screen, the fractional amounts emerging from the screen in both the forward and backward directions.

Since the x-ray intensity decreases exponentially with depth of penetration, light production also follows an exponential law, being highest at the back where the x-rays are incident. Consider the element of thickness $dx$ located at a distance $x$ from the back of the screen whose thickness is $\epsilon$ (Fig. 8). This element is between two layers, of finite thickness $x$ (layer 1) and $\epsilon-x$ (layer 2), respectively. Of the light $dL$ generated in the element, $\frac{1}{2}$ starts out toward layer 2, while $\frac{1}{2}$ starts toward layer 1. These are subject to multiple reflections as shown in Fig. 8, the lower half being concerned with the light starting toward 2 and the upper with that starting toward 1. The fraction $f$ out the front due to $dL$ is given by the sum of the contributions as before, and

$$f=\frac{1}{2}\,\frac{T_2(1+R_1)}{1-R_1R_2}. \tag{2.5}$$

The coefficients $R_1$, $R_2$ and $T_2$ are, of course, functions of $x$ and $\epsilon-x$ as in formulas (2.3) and (2.4). It is now necessary to multiply $f$ by the light produced in the element $dx$ and integrate over $x$ from 0 to $\epsilon$ to obtain the total light out.

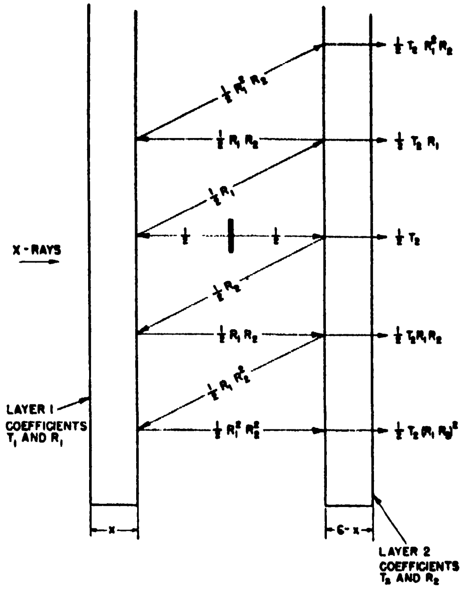


FIG. 8. Schematic representation of light flow from a luminous element within the screen.

If the absorption coefficient of the x-rays is $\mu$, we have by definition that the x-rays absorbed in a layer $dx=I\mu dx$ where $I$ is the intensity of x-rays at the point $x$. Let $\rho$ = conversion efficiency of x-rays into light; the light produced in $dx$ is then $\rho I\mu dx$. But $I=I_0e^{-\mu x}$, where $I_0$ is the incident x-ray intensity, and $dL=\rho\mu I_0e^{-\mu x}dx$ (2.6)

The forward light yield is then:

$$L=\frac{\rho\mu I_0}{2}\int_0^{\epsilon}e^{-\mu x}\frac{T_{(\epsilon-x)}(1+R_{(x)})}{(1-R_{(x)}R_{(\epsilon-x)})}dx. \tag{2.7}$$

Before performing this integration we shall find it very convenient to derive a relation between the transmission coefficient of the screen as a whole ($T_\epsilon$) and the coefficients of its parts. In Fig. 9 is shown the schematic diagram from which we find:

$$\begin{aligned}T_\epsilon&=T_1T_2(1+R_1R_2+(R_1R_2)^2+\cdots),\\ T_\epsilon&=T_1T_2/(1-R_1R_2).\end{aligned} \tag{2.8}$$

Substituting in formula (2.5) we get

$$f\doteq\frac{T_\epsilon}{2}\left(\frac{1}{T_1}+\frac{R_1}{T_1}\right). \tag{2.9}$$

$T_\epsilon$ is a constant independent of $x$, and from 2.3 and 2.4 we have $1/T_1 = \cosh px + (\alpha/p)\sinh px$, $R_1/T_1 = (k/p)\sinh px$. Formula 2.7 then becomes:

$$L = \frac{p\mu I_0 T_\epsilon}{2}\int_0^\epsilon e^{-\mu x} \times\left(\cosh px + \frac{\alpha+k}{p}\sinh px\right)dx. \quad (2.10)$$

In preference to an expression for light out itself, we shall divide $L$ by the total light generated to obtain the fraction $F_F$ which is transmitted out the front of the screen. The total light produced is:

$$pI_0(1-e^{-\mu\epsilon})$$

and

$$F_F = \frac{\mu}{(1-e^{-\mu\epsilon})[\cosh p\epsilon + (\alpha/p)\sinh p\epsilon]} \times\int_0^\epsilon e^{-\mu x}\left(\cosh px + \frac{\alpha+k}{p}\sinh px\right)dx. \quad (2.11)$$

Carrying out the integration we obtain the final expression:

$$F_F = \frac{\mu\{(\mu+\alpha+k) - e^{-\mu\epsilon}[(\mu+\alpha+k)\cosh p\epsilon + ([\mu(\alpha+k)/p]+p)\sinh p\epsilon]\}}{2(1-e^{-\mu\epsilon})(\mu^2-p^2)(\cosh p\epsilon + (\alpha/p)\sinh p\epsilon)}. \quad (2.12)$$

In an exactly similar manner the formula for the fraction out the back of the screen is found;

$$F_B = \frac{\mu\{(\mu-\alpha-k)(\cosh p\epsilon - e^{-\mu\epsilon}) + ([\mu(\alpha+k)/p]-p)\sinh p\epsilon\}}{2(1-e^{-\mu\epsilon})(\mu^2-p^2)(\cosh p\epsilon + (\alpha/p)\sinh p\epsilon)}. \quad (2.13)$$

Graphs of these functions are given in Figs. 10 and 11 for various values of $\mu$ and $\epsilon$, using $k = 5.55$/layer and $p = 0.30$/layer. Values of the thickness $\epsilon$ are in terms of one-layer units, and $\mu$, the x-ray absorption coefficient is per layer. It is apparent that for zero thickness of screen half of the light is transmitted in each direction regardless of $\mu$. At the limit $\mu \rightarrow 0$ light is produced uniformly throughout the screen, and for thicker screens equal amounts of somewhat less than 50 percent are emitted in each direction, the remainder being absorbed in the screen. For thick screens and high $\mu$, the fraction emitted toward the back is much more than 50 percent, while the light forward is much less. For $\mu = \infty$, all of the light is produced in a layer of zero thickness at the back. Half of the light is immediately emitted backwards and a fraction $R(\epsilon)$ of the remaining half is reflected in addition. The fraction out the front is the transmission coefficient times the half originally directed toward the screen. Thus all the curves of Figs. 10 and 11 approach different limits as $\mu \rightarrow \infty$.

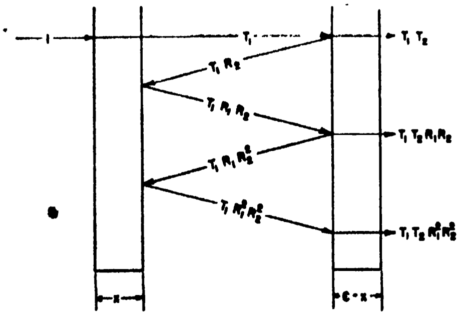


FIG. 9. Light flow diagram for derivation of formula relating partial and total transmission coefficients.

While these curves show wide variations in yields, actually the ranges of $\mu$ and $\epsilon$ usually encountered in practice are such that the fractions forward and back are fairly close to 50 percent. At 80 kv constant potential, $\mu$ is approximately 0.25 per layer; for a screen of $\epsilon = 1$ (Eastman Ultra-Speed front) the fraction forward is 47.6 percent, the fraction back 51.2 percent. At higher voltages little error is introduced by assuming 50 percent each way unless the screen is abnormally thick. At the lower voltages, $\mu$ rises rapidly, attaining a value of 2.2 at 20 kv and 3 at 17.4 kv. Here the fractions out the front and back are markedly affected by the operating voltage.

### Effect of Mounting Card

In practice a screen is never used without a backing of some sort. Usually this is made white, with the intent of recovering the light emitted

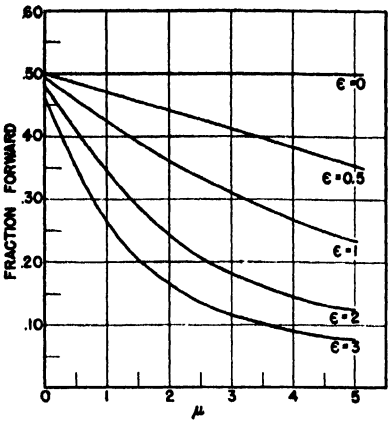


FIG. 10. Fraction of light generated which escapes from the front of the screen as a function of screen thickness ($\epsilon$) and x-ray absorption coefficient ($\mu$).

in the backward direction. The efficacy of this procedure will now be examined. Figure 12 shows schematically a layer of fluorescent material backed by a card of reflection coefficient $R_0$.

The ratio of the light returned to the light emitted backward from the screen is defined as $E$.

$$E = R_0T(1 + RR_0 + (RR_0)^2 + \cdots, \qquad (2.14)$$

$$E = \frac{R_0T}{1 - RR_0}.$$

By way of example, if a screen of thickness $\epsilon = 1$, $R = 0.844$ and $T = 0.15$, is backed by a card of reflection coefficient $R_0 = 0.80$, one finds $E = 0.37$, so of all light which was previously lost out the back of the screen, only 37 percent is recovered by using a backing card of 80 percent reflectivity. This is the result of the "whiteness" of the calcium tungstate; if the card is not of superior reflecting power, it is of small value. Some of this loss is in the layer itself, but the major portion goes into the card. The card upon which the Eastman screen was mounted had a diffuse reflection coefficient of 0.84. Table II illustrates the role played by this mounting card in the light losses. These data were computed for 80 kv, using the constants $k = 6.0$ and $p = 0.33$. At higher voltages very little change takes place in the light distribution, and the tabulated values will remain essentially constant. At lower voltages the losses will increase considerably, reaching a limit at $\mu = 0$ of 52 percent for the thin screen and 74 percent for the thick screen.

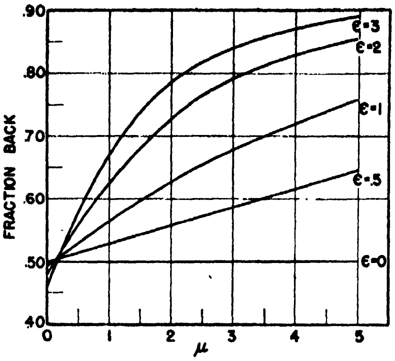


FIG. 11. Fraction of light generated which escapes from the back of the screen as a function of screen thickness ($\epsilon$) and x-ray absorption coefficient ($\mu$).

The increase in light to be expected from the use of a card has been plotted in Fig. 13. At first glance these numbers may seem quite high; however, it must be remembered that in the vast majority of cases $\mu$ is 0.5 or smaller, so that the increase will seldom exceed 40 percent.

Two experimental checks of the above formulas were made. In the first the light from a fluorescent screen was measured with and without a backing card. In the second, two screens of different thickness were measured and the expected relative light output calculated on the basis of the above formulas. Both tests checked calculated values within 3 percent.

So far we have not discussed the variation with thickness of the amount of x-rays absorbed. This, of course, is the most important factor in determining light output for thin screens. With increasing screen thickness the x-rays are more and more nearly completely absorbed; in addi-

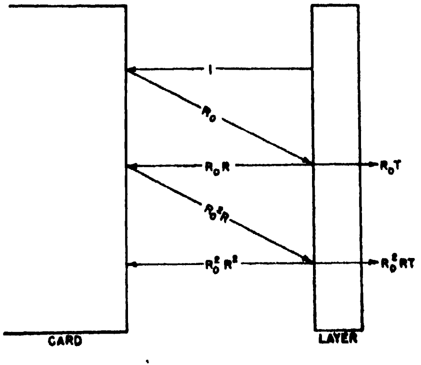


FIG. 12. Light flow diagram for derivation of effect of reflecting card.

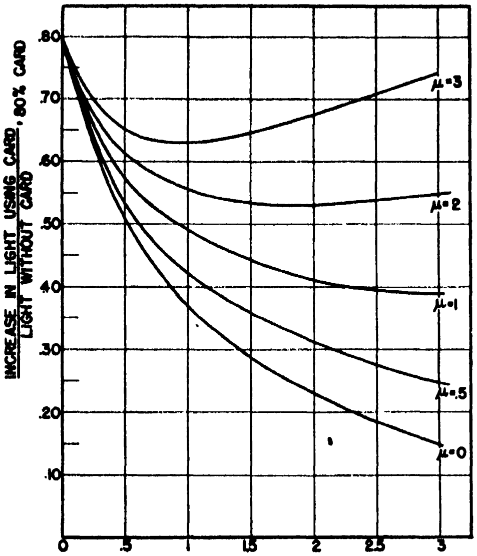


FIG. 13. Increase in light to be gained by using a backing card with 80 percent reflectance, as a function of screen thickness ($\epsilon$) and x-ray absorption coefficient ($\mu$).

tion, the effectiveness of the card backing decreases and losses in the phosphor and binder begin to gain importance. Thus the light output reaches a maximum at a thickness which will depend upon the absorption coefficient (or voltage) of the x-rays. In Figs. 14 and 15 is shown the utilization factor $U$, defined as the ratio of the light obtained from the front of the screen to the total light which would have been generated had all the x-rays been absorbed. The curves of Fig. 14 were calculated for a screen without mounting card, those of Fig. 15 with a mounting card having 100 percent reflection. It will be seen that in practical cases the utilization is seldom over 30 percent.

In the case of a "back" screen one is interested in the light out the back, i.e., toward the x-ray tube. This quantity increases with the screen thickness, but the gain is very slight beyond a certain point. A thick layer of $CaWO_4$ is much whiter than any available cardboard; for a simple example we can just set the thickness of the back screen at infinity and inquire how the utilization factor varies with $\mu$. If $\epsilon$ is infinite, $F_B$ in formula 2.13 is the utilization factor, since all the x-rays are absorbed. The formula reduces to:

$$U_B = [(p+\alpha+k)/2(p+\alpha)](\mu/\mu+p). \quad (2.14)$$

The factor is plotted in Fig. 16 using values of $p$ and $\alpha$ for $CaWO_4$ as before. It is seen that even with an infinitely thick screen the utilization is not very high for $\mu$ less than 0.3, i.e., for x-ray voltages above 80 kv.

In the design of screens for radiographic purposes an all-important consideration is the resolving power of the screen-film combination. In general, the thickness of the screen is largely determined by the stringency of this requirement. No treatment of resolving power is attempted here, because the primary purpose of the investigation is to measure the true efficiency of the $CaWO_4$ screen. The curves in the preceding portion show accurately the light produced as the screen thickness and x-ray absorption coefficient vary, but they form a basis for selection of screen thickness only on the assumption that resolving power is to be ignored. For this reason it is of little value to extend the treatment to

TABLE II. Light losses in two typical x-ray intensifying screens.

| | Thin screen (1 layer) .045 gm/cm² $CaWO_4$ | Thick screen 2.45 layer .109 gm/cm² $CaWO_4$ |
|---|---|---|
| Light lost in phosphor and binder | 2.2% | 9.6% |
| Light lost in mounting card | 28.7% | 36.0% |
| Light yield (total) | 69.1% | 54.4% |

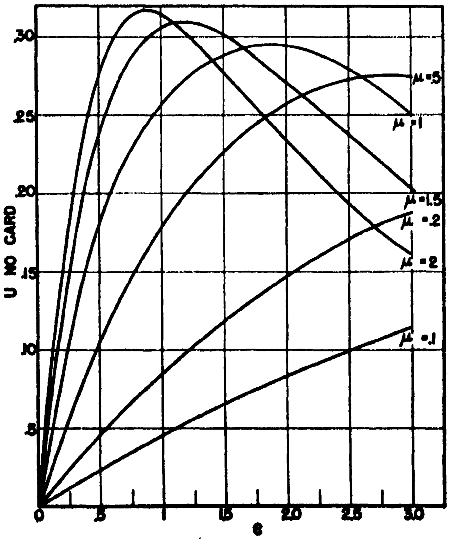


FIG. 14. X-ray utilization factors for a screen with a non-reflecting backing.

other cases, such as a front and back screen combination. The information presented here is sufficient to calculate the light yield of any such combination, given the absorption coefficient, thicknesses, and mounting card reflectances of the screens used.

It should be pointed out here that if the film used with a screen is itself reflecting, all of the light will not be absorbed in the film. Measured white-light reflectances of x-ray screen film are high enough to cause errors of 8 percent with an Ultra-Speed front screen; however, it is quite possible that the reflectance is different for the blue and ultraviolet light of the screens.

## III. EFFICIENCY OF FLUORESCENCE

To obtain the x-ray energy absorbed in the screen, it is necessary to have a knowledge of the energy spectrum of the incident radiation and of the absorption of energy from this spectrum by the screen.

Information as to the exact continuous spectrum of x-ray radiation from a thick target is not available in the literature. Curves originally published by Ulrey[1] have been reprinted in all the texts for 25 years in spite of the fact that they were known to be in serious error at the time of publication. Our use of these curves in combination with the known absorption coefficients of $CaWO_4$ led to obvious and large discrepancies with the measured absorption of the screen, and an independent determination of the x-ray spectrum seemed desirable. Figure 17 shows the energy distribution from the x-ray tube used by us, as determined by Mr. R. L. Longini with a method to be published by him at a later date. This spectrum applies to the tube run at 80 kvc with $\frac{1}{2}$ mm copper and 1 mm aluminum filter. On the same figure is given the curve of energy absorbed by the screen, calculated from the known total absorption coefficients. The integrated areas under these curves, when converted to ionizing power, agreed with our measured absorption value.

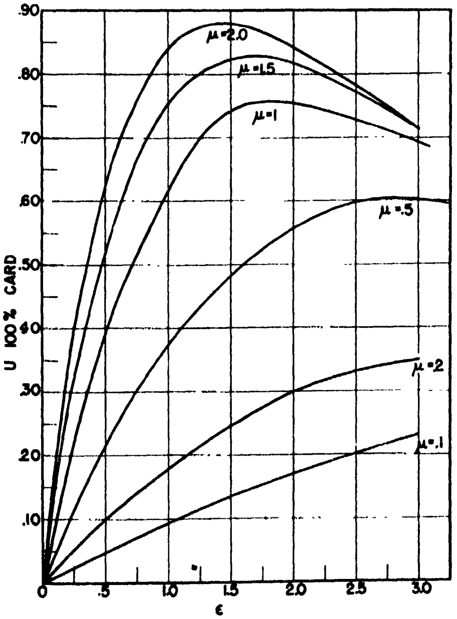


FIG. 15. X-ray utilization factors for a screen with a perfectly reflecting back.

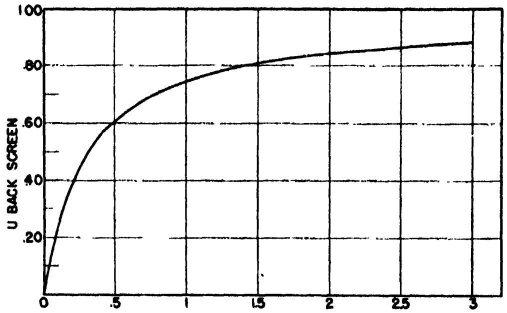


FIG. 16. X-ray utilization factor for a very thick screen used as a back screen, plotted against μ.

All x-radiation which leaves the screen is considered not to be part of the input for the purpose of our computation of conversion efficiency. Our measurement of the undeflected x-ray beam leaving the screen fails to include the scattered and fluorescent x-radiation which leaves under large angles. The former can be shown to be negligible at the voltages used; the latter is important only for energies above the *K*-absorption edge of tungsten. Of the 299 ergs absorbed in the region to the left of the *K*-edge in Fig. 17, absorption by *K*-electron ejection includes 190 ergs which will be re-emitted as *K*-radiation. Of this, about 5 percent is converted into Auger electrons and 26.5 percent of the remainder is reabsorbed in our particular thickness of screen, so that 133 ergs escaped as x-rays. Thus the total x-ray input must be reduced from 714 ergs to 581 ergs.

---

[1] C. T. Ulrey, Phys. Rev. 11, 401 (1918).

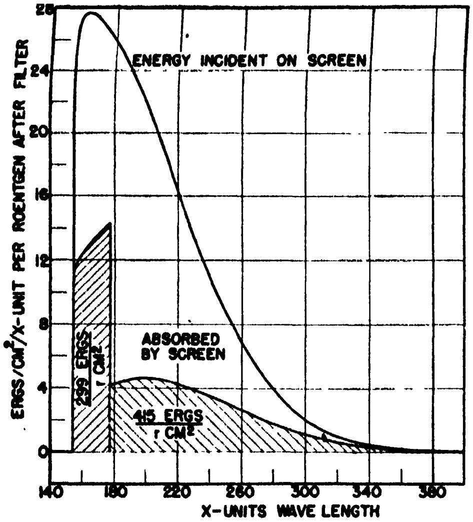


FIG. 17. Spectral energy distributions of x-rays incident on the screen and x-rays absorbed by the screen.

We are now in a position to calculate the efficiency of calcium tungstate. From Part I we have the light out of one side as 2.34 ergs/cm$^2$/sec with an incident radiation of 8.92 r/min. At 581 ergs/cm$^2$ r we have then 86.2 ergs/cm$^2$/sec absorbed in the screen. Thus the "gross" efficiency is 2.72 percent. From Part II, however, we find that the screen emits only 54.5 percent of the light generated, and the intrinsic efficiency of the calcium tungstate is finally found to be 5.0 percent.

This efficiency may be expected to vary only slightly with the wave-length of the x-radiation involved, since one electron ejected by an x-photon excites many fluorescent centers and dissipates its energy gradually.

It is of interest to compute the number of fluorescent centers activated by a single absorbed x-photon. Taking the average wave-length of the absorbed x-rays as 0.22 angstroms, and the average wave-length of the light produced as 4400 angstroms, we see that energetically it would be possible to get 20,000 light quanta from one x-photon. Since the material is only 5 percent efficient, however, on the average each x-photon absorbed produces 1000 light quanta.

# Pressure and Oil Flow in Oil-filled Cables at Load Variations

HANS LOTTRUP KNUDSEN
*Holstebro, Denmark*

(Received September 23, 1946)

This paper presents a method for calculating pressure and oil flow in a section of an oil-filled cable at load variations with due regard to both the hydraulic resistance of the oil canals and the elasticity of the sheath and the pressure armor, if any, and the compressibility of the oil. The present article restricts itself to cable sections terminated with oil reservoirs in which the pressure is constant; in a forthcoming paper this restriction will be dropped. Formulas are given for pressure, oil current, and amount of oil which flows through a cross section of the cable for an arbitrarily prescribed oil expansion per unit length. These formulas are applied to the following oil expansion functions: $S_0\cdot 1$, $d/dt(N_0\cdot 1)$, $e^{-t/T}\cdot 1$, and $1/\sqrt{t}\cdot 1$, which represent the oil expansion or terms occurring in the series giving the oil expansion in some important cases. Generalized functions are plotted for the pressure at the midpoint, the oil current at the endpoints, and the amount of oil which passes the endpoints for the above mentioned oil expansion functions. Using these generalized functions, the calculation of pressure and oil flow, with the simplifying assumptions made here, is reduced to elementary operations. Numerical examples illustrate the use of the given formulas and curves.

---

## LIST OF SYMBOLS

| | | |
|---|---|---|
| $a$ | cm$^2$ | Area of oil canal. |
| $A$ | cm$^2$ | Oil volume per unit length. |
| $c$ | kg$^{-1}$ cm$^4$ | Capacitance per unit length. |
| $c_1$ | kg$^{-1}$ cm$^4$ | Capacitance per unit length corresponding to the elasticity of sheath and pressure armor. |
| $c_2$ | kg$^{-1}$ cm$^4$ | Capacitance per unit length corresponding to the compressibility of oil. |
| $C$ | kg$^{-1}$ cm$^5$ | Capacitance of the entire cable. |
| $d$ | cm | Diameter of oil canal in single-core cable. |
| $D$ | cm | Diameter of cable under sheath. |
| $E$ | kg cm$^{-2}$ | Modulus of elasticity for the sheath material. |
| $f$ | cm | Perimeter of the conductors in which short-circuit current flows. |
| $F(x)$ | 1 | Function defined in (54). |
| $I(x, t)$ | cm$^3$ sec.$^{-1}$ | Current of oil. |
| $k$ | kg$^{-1}$ cm$^2$ | Specific compressibility of oil. |
| $K_0$ | cm$^2$ sec.$^{-\frac{1}{2}}$ | Constant factor in an expression for the oil expansion. |
| $l$ | cm | Arbitrarily defined length in cable cross section. |
| $L$ | cm | Length of cable. |
| $M(x, t)$ | cm$^3$ | Amount of oil which flows through a cross section of the cable. |
| $N_0$ | cm$^2$ | Instantaneous oil expansion per unit length. |
| $p$ | °C cm $W^{-1}$ | Specific thermal resistance for insulation. |
| $P(t)$ | cm$^{-1}$ $W$ | Effect transferred from conductors to insulation per unit length. |
| $q$ | °C$^{-1}$ cm$^{-3}$ $W$ sec. | Specific heat of insulation. |
| $r$ | kg cm$^{-6}$ sec. | Hydraulic resistance in cable per unit length. |
| $R$ | kg cm$^{-5}$ sec. | Hydraulic resistance of entire cable. |
| $S(t)$ | cm$^2$ sec.$^{-1}$ | Oil expansion per unit time per unit length. |
| $S_0$ | cm$^2$ sec.$^{-1}$ | Constant factor in expression for oil expansion. |
| $t$ | sec. | Time. |
| $T$ | sec. | Time constant occurring in the expression for oil expansion. |
| $V(x, t)$ | kg cm$^{-2}$ | Pressure. |
| $\beta$ | °C$^{-1}$ | Volumetric expansion coefficient of oil. |
| $\beta'$ | °C$^{-1}$ | Volumetric expansion coefficient of insulation. |
| $\gamma$ | g cm$^{-3}$ | Density of oil. |
| $\delta$ | cm | Sheath thickness. |
| $\vartheta(x, t)$ | °C | Temperature in insulation. |
| $\vartheta_0$ | °C | Temperature of conductors after short circuit. |
| $\nu$ | centistokes | Viscosity of oil. |

## PRESSURE AND FLOW QUALITATIVELY REGARDED

IN the calculation of oil-filled cable systems it is important to have a method for computing the variations of the pressure for a certain system under the most severe load variations to which it can be subjected in service. These commonly consist of (1) switching off the load and (2) short circuiting. In the first case the oil in the cable is cooled and consequently the pressure decreases below the normal value until an amount of oil equaling the contraction of the oil has flowed into the cable. In consequence of the diminished pressure the electrical strength of the cable may decrease so much that breakdown occurs. On the other hand, in the case of a short circuit the oil expands rapidly because of the rise in temperature, and the pressure in the cable will for some time increase and later decrease as the surplus of the oil flows towards the oil reservoirs of the system. In the case of a short circuit the pressure in the cable may increase so much that there is a risk of the sheath and the pressure armor being destroyed.

All calculations of the pressure must be based upon the variations of the temperature in the cable with the given load variations. Several methods of computing the variations of the temperature and the oil expansion at load variations in cables placed under different circumstances have been described in the literature. Some of these methods will be mentioned later.

When the oil expansion has been calculated, the variations of the pressure and the oil flow will be investigated. The dependence of the temperature on load variations has been treated thoroughly in the literature. That is not the case with the pressure and the oil flow.

In many cases the increase or decrease in pressure for the greatest expansion of oil per unit time at load variations has been calculated under the assumption of stationary oil flow in the cable. For oil cable systems with different types of oil reservoirs, Emanueli[1] has developed a theory for the calculation of pressure and oil flow, in which the pressure drop along the cable is calculated as if the oil flow were stationary. Shanklin and Buller[2] make the same assumption in their calculations of the pressure drop.

H. Spanne, chief engineer at Sieverts Kabelverk, Sundbyberg, Sweden, however, demonstrated about ten years ago in a factory bulletin, that the elasticity of the sheath and of the pressure armor, if any, and the compressibility of the oil in an oil-filled cable greatly limit the pressure variations occurring at load changes. Spanne has also made an estimate of the pressure variations.

In the present article a method has been developed for calculating the variations of pressure and the oil flow in an oil-filled cable with reservoirs at both ends, for an oil expansion given as a function of the time, with due regard to the flow resistance in the cable as well as to the elasticity of the sheath and the presssure armor, if any, and to the compressibility of the oil. Only those cases in which the expansion of oil per unit time and per unit length is the same at every point of the cable have been treated here.

## SIMPLIFYING ASSUMPTIONS

The following assumptions are the most important of those made in this paper:

1. All deformations in the cable are purely elastic.
2. No force is needed to accelerate the oil in the canals.
3. The viscosity of the oil in the oil canals is constant.
4. The pressure in the oil reservoirs is constant.

Assumptions 1 and 2 are approximately satisfied in most practical oil-filled cables. As the viscosity of the oil varies much with the temperature, 3, however, is not well satisfied. The difficulties involved in solving the problem are, however, greatly increased if assumption 3 is not made. Assumption 4 is well satisfied for cable systems with oil reservoirs consisting of oil tanks with atmospheric pressure at the surface of the oil. However, assumption 4 is often rather well satisfied in other types of oil reservoirs, too, as the oil expansion in the cable for a load variation often is small compared with the oil expansion corresponding to the total temperature variation for which the system has been constructed. (The total temperature includes, for instance; the seasonal change of temperature.) This statement applies especially in the case of a short circuit in which the temperature of the conductors may certainly be rather high, but in which the total developed heat energy, and consequently also the oil expansion, often are small.

An article is being prepared, in which assumption 4 is not made, due regard on the contrary being given to the characteristics of different oil reservoirs.

---

[1] L. Emanueli, Trans. A.I.E.E. 47, 186 (1928).

[2] G. B. Shanklin and F. H. Buller, Trans. A.I.E.E. 50, 1411 (1931).

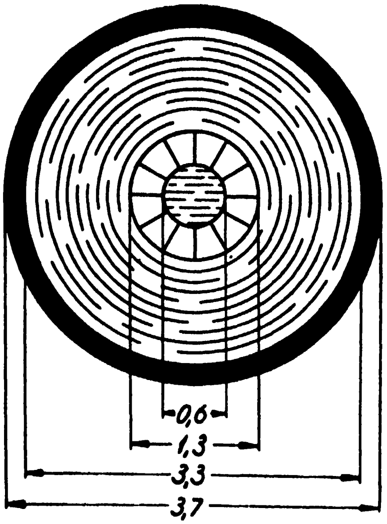


FIG. 1. Single core oil-filled cable, $77/\sqrt{3}kV$, 1×95 mm.[2]

## MATHEMATICAL FORMULATION

The mathematical expressions for the flow in an oil-filled cable depend on how the flowing medium is defined. This, for example, can be defined as the mass of the oil, or as the volume of the oil. Here we shall base the calculations on a third and more convenient definition, defining the flowing medium as the volume of the oil reduced to constant pressure.

The amount of medium per unit length at a certain point of the cable will then increase proportionally to the pressure because of the elasticity of the sheath and of the pressure armor, if any, as well as of the compressibility of the oil. We define the capacitance per unit length, $c$, as the ratio of the increase in amount of medium per unit length to the increase of pressure. The capacitance $c$ is then composed of two parts, one part $c_1$, arising from the elasticity of the sheath and the pressure armor, and another part $c_2$, arising from the compressibility of the oil.

When no pressure armor is present, the capacitance $c_1$ for a cable with a sheath of thickness $\delta$, diameter $D$, and modulus of elasticity $E$ is

$$c_1=\frac{\pi D^3}{4\delta E}. \tag{1}$$

It is a trivial matter to develop the corresponding formula for cables with a pressure armor.

$c_2$ is calculated from the compressibility of the oil, $k$, and the amount of oil per unit length, $A$.

$$c_2=kA. \tag{2}$$

$c$ is thus

$$c=c_1+c_2. \tag{3}$$

The current of the medium is defined as the amount of oil which passes a section of the cable per unit time. The resistance per unit length is defined as the ratio of the pressure drop per unit length to the current.

The flow at the moderate currents in question will be laminar, and $r$ can consequently be expressed as follows,[3]

$$r=B\frac{\gamma\nu}{l^4}, \tag{4}$$

where $\gamma$ is the density, $\nu$ the kinematic viscosity of the oil, $l$ a length which increases proportional to the linear dimensions of the section of the oil canal, and $B$ a constant which merely depends on the shape of the oil canal and the way in which $l$ has been defined. If $l$ stands for the inner diameter of oil canals with a circular section one gets

$$B=\frac{128}{9.81\cdot\pi}\cdot 10^{-7}=4.15\cdot 10^{-7}.$$

For a cable with a length $L$ the total capacitance is $C=Lc$ and the total resistance $R=Lr$. The time constant $RC$ for the oil flow in the cable is a measure of the rate at which disturbances in the state of equilibrium of the cable are smoothed. In order to give an idea of the order of magnitude of the time constant it has been computed below in a practical case for two different temperatures.

### Example 1

The time constant at the temperatures 20°C and 40°C shall be calculated for an oil-filled cable $77\sqrt{3}kV$, 1×95 mm², without pressure armor and with the principal dimensions as shown in Fig. 1.

$D=3.3$ cm,

$d=0.6$ cm,

---

[3] *Handbuch der Physik* (1927), Vol. 7, p. 103.

$$\delta = 0.2\ \text{cm},$$

$$A = 3.3\ \text{cm}^2,$$

$$L = 10^5\ \text{cm},$$

$$E = 1.7\cdot 10^5\ \text{kg cm}^{-2},$$

$$k = 8\cdot 10^{-5}\ \text{kg cm}^2,$$

$$\gamma\nu = \begin{cases} 44.8\ \text{centipoise for } 20^\circ\text{C} \\ 16.4\ \text{centipoise for } 40^\circ\text{C}, \end{cases}$$

$$c_1 = \frac{\pi\cdot 3.3^3}{4\cdot 0.2\cdot 1.7\cdot 10^5} = 8.3\cdot 10^{-4}\ \text{kg}^{-1}\ \text{cm}^4,$$

$$c_2 = 8\cdot 10^{-5}\cdot 3.3 = 2.6\cdot 10^{-4}\ \text{kg}^{-1}\ \text{cm}^4,$$

$$c = 1.09\cdot 10^{-3}\ \text{kg}^{-1}\ \text{cm}^4,$$

$$r = \frac{128}{9.81\cdot\pi}\cdot 10^{-7}\cdot\frac{\gamma\nu}{0.6^4} = 3.2\cdot 10^{-6}\cdot\gamma\nu$$

$$= \begin{cases} 1.43\cdot 10^{-4}\ \text{kg cm}^{-6}\ \text{sec. for } 20^\circ\text{C} \\ 5.3\cdot 10^{-5}\ \text{kg cm}^{-6}\ \text{sec. for } 40^\circ\text{C}, \end{cases}$$

$$RC = rcL^2 = \begin{cases} (1.43\cdot 10^{-4})\cdot(1.09\cdot 10^{-3})\cdot 10^{10} \\ \qquad = 1560\ \text{sec. for } 20^\circ\text{C}, \\ (5.3\cdot 10^{-5})\cdot(1.09\cdot 10^{-3})\cdot 10^{10} \\ \qquad = 580\ \text{sec. for } 40^\circ\text{C}. \end{cases}$$

This example shows that the time constant for the oil flow is not negligible compared to the intervals of time commonly necessary for smoothing the temperature in a cable under load variations. This means that the elasticity of the sheath and the compressibility of the oil will limit the pressure variations in the cable with changes in load.

The differential equations and the boundary conditions for the oil flow will now be given. Because we have assumed that the cable has constant parameters, without making the treatment of the problem less general we can suppose that the static pressure in the cable and consequently also the pressure at the endpoints of the cable are zero at any moment. Over the dynamic pressure distribution, calculated according to this assumption, we can thus superimpose the static pressure distribution in the cable. In this paper only the dynamic pressure distribution is considered. For a given oil expansion per unit length of the cable per unit time, $S(t)$, we shall calculate as functions of time and position the pressure $V$, the current $I$, and the amount of oil which flows through a cross section of the cable, $M$.

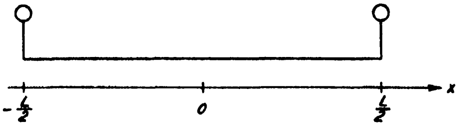


FIG. 2. Coordinate system along a cable section terminated with oil reservoirs.

A coordinate axis is placed along the cable with the origin at the midpoint (see Fig. 2). $V$ and $I$ then satisfy the following partial differential equations.

$$c\frac{\partial V}{\partial t} = \frac{1}{r}\frac{\partial^2 V}{\partial x^2} + S, \tag{5}$$

$$I = -\frac{1}{r}\frac{\partial V}{\partial x}, \tag{6}$$

$$V = 0 \text{ for } x = \begin{cases} -L/2 \\ L/2 \end{cases} \quad 0 < t < \infty, \tag{7}$$

$$V = 0 \text{ for } t < 0, \quad -L/2 < x < L/2. \tag{8}$$

Equations formally identical with the equations given above are obtained for several problems within other domains of physics. Suppose we have a metallic wire, isolated against heat, the length being $L$, with the heat conduction resistance per unit length $r$, the heat capacitance per unit length $c$, and with the endpoints at the temperature zero. If an electrical current producing the effect per unit length $S(t)$ is flowing in the wire, the temperature $V$ and the heat current $I$ satisfy Eqs. (5)–(8).

There also exists an electrical analogy to the problem in question. If current sources of the strength per unit length $S(t)$ are supposed homogeneously distributed along a telegraph cable of length $L$, without inductance and leakage but with resistance per unit length $r$, capacitance per unit length $c$ (Thomson cable), and with the ends short-circuited, the potential $V$ and the current $I$ will satisfy Eqs. (5)–(8).

In the oil-flow problem the mass of the oil corresponds to the inductance of the cable in the electrical analogy. The mass of the oil has no corresponding meaning in the problem of heat conduction.

It is convenient to solve first the problem given by Eqs. (5)–(8) for $S(t)=S_0\cdot 1$, where $S_0$ is a constant.

### PRESSURE AND FLOW WITH CONSTANT OIL EXPANSION $S_0\cdot 1$

The thermal analogy to this problem has been treated by Straneo,[4] Schaufelberger,[5] and Weinreich[6] for use in measuring the electrical and thermal constants of metallic wires. The authors mentioned above solve the problem by use of eigenfunctions following classical methods.

The problem, however, is very conveniently solved by Heaviside's operational calculus. The appropriate method has been described e.g., by Dahr[7] in his treatment of the telegraphist's equation for transmission lines in which electromotive forces are induced. (See also Jeffrey's treatment of related problems.[8]) For convenient reference the application of the method on the present problem is briefly outlined below.

Introducing the operators

$$\mathfrak{L}V(x,t)=v(x,p),\tag{9}$$

$$\mathfrak{L}I(x,t)=i(x,p),\tag{10}$$

Eqs. (5)–(8) are transformed into the following equations:

$$cpv=-\frac{1}{r}\frac{d^2v}{dx^2}+S_0,\tag{11}$$

$$i=-\frac{1}{r}\frac{dv}{dx},\tag{12}$$

$$v=0\ \text{for}\ x=\begin{cases}-L/2\\ L/2.\end{cases}\tag{13}$$

First the more simple problem of a point source 1 acting at the point $x=\xi$ is solved. For the pressure and the current in this problem the differential equations that are obtained from (11) and (12) by setting $S_0=0$ are valid, while to the boundary condition (13) are added the conditions

$$v(\xi-0)=v(\xi+0),\tag{14}$$

$$i(\xi-0)+1=i(\xi+0).\tag{15}$$

As the solution of this problem one gets the following expression for the pressure, which in this special case will be called $w(x,\xi)$,

$$w(x,\xi)=\begin{cases}-rL\dfrac{\sinh\left(-\frac{1}{2}+\frac{\xi}{L}\right)(pRC)^{\frac{1}{2}}\sinh\left(\frac{1}{2}+\frac{x}{L}\right)(pRC)^{\frac{1}{2}}}{(pRC)^{\frac{1}{2}}\sinh(pRC)^{\frac{1}{2}}}=w_1(x,\xi)\ \text{for}\ -\dfrac{L}{2}<x<\xi\\[2ex] -rL\dfrac{\sinh\left(\frac{1}{2}+\frac{\xi}{L}\right)(pRC)^{\frac{1}{2}}\sinh\left(-\frac{1}{2}+\frac{x}{L}\right)(pRC)^{\frac{1}{2}}}{(pRC)^{\frac{1}{2}}\sinh(pRC)^{\frac{1}{2}}}=w_2(x,\xi)\ \text{for}\ \xi<x<\dfrac{L}{2}.\end{cases}\tag{16}$$

The expression for $v$, in the case of a constant, homogenously distributed source, is now obtained by integration of $S_0w$ over the entire length of the cable

$$v(x,p)=S_0\int_{-L/2}^{L/2}w(x,\xi)d\xi=S_0\left[\int_{-L/2}^{x}w_2(x,\xi)d\xi+\int_{x}^{L/2}w_1(x,\xi)d\xi\right]$$

$$=S_0rL^2\left[\frac{1}{pRC}-\frac{\sinh\left(\frac{1}{2}+\frac{x}{L}\right)(pRC)^{\frac{1}{2}}-\sinh\left(-\frac{1}{2}+\frac{x}{L}\right)(pRC)^{\frac{1}{2}}}{pRC\sinh(pRC)^{\frac{1}{2}}}\right].\tag{17}$$

---

[4] P. Straneo, Roma. Rend. Acc. dei Lincei (Ser. 5) 7, 206 (1898).

[5] W. Schaufelberger, Ann. d. Physik 7, 589 (1902).

[6] W. Weinreich, Zs. Math. (Leipzig) 63, 1 (1914).

[7] K. Dahr, *A Course of Integrational and Operational Calculus* (V. Pettersons Bokindustriaktiebolag, Stockholm, 1935), p. 44.

[8] H. Jeffreys, "Operational Methods in Mathematical Physics," Cambridge Tracts in Math. and Math. Ph. 23, 49 (1931).

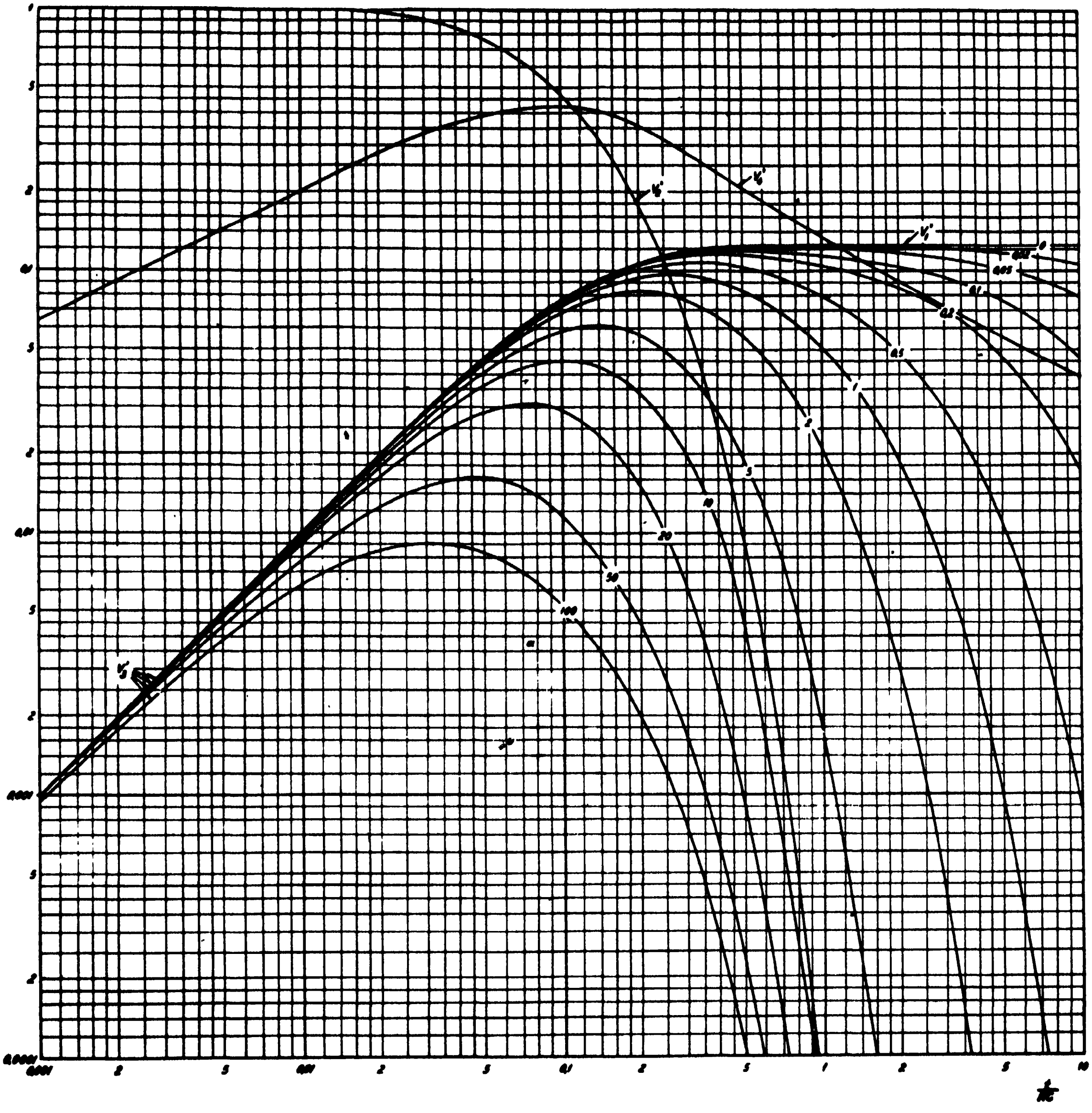


FIG. 3. Generalized functions used in the calculation of the pressure at the midpoint of a cable section with a time constant for the oil flow *RC*.

$V_1'(t/RC)$ applies in the case of an oil expansion $S(t)=1\cdot 1$.
$V_2'(t/RC)$ applies in the case of an oil expansion $S(t)=(d/dt)(1\cdot 1)$.
$V_3'(t/RC)$ applies in the case of an oil expansion $S(t)=e^{-\alpha t/RC}\cdot 1$.
$V_4'(t/RC)$ applies in the case of an oil expansion $S(t)=1/(t)^{\frac{1}{2}}\cdot 1$.

The time function for operators of this type can be obtained either by expanding the operator in an infinite series of terms containing exponential functions and transforming this series, term by term, or by applying Heaviside's expansion theorem. The first method corresponds to expressing the solution in reflected waves as demonstrated e.g., by Wagner,[9] whereas by the last method the solution is expressed in eigenfunctions. Here we shall use the last method.

According to Heaviside's expansion theorem the time function $H(t)$ that corresponds to an

---

[9] K. W. Wagner, Operatorenrechnung (Johann Ambrosius Barth Verlag, Leipzig, 1940), p. 196.

operator $f(p)/g(p)$, where $f(p)$ and $g(p)$ are functions of $p$, which satisfy certain conditions, and where $g(p)=0$ has the simple roots $p_n$, $n=1, 2, 3, \cdots$, is

$$H(t)=\mathfrak{L}^{-1}\frac{f(p)}{g(p)}=\frac{f(0)}{g(0)}+\sum\frac{f(p_n)e^{p_nt}}{p_ng'(p_n)}. \tag{18}$$

If we write the operator $v$ (17) as follows

$$v=\frac{f(p)}{g(p)}, \tag{19}$$

where

$$f(p)=S_0rL^2\frac{1}{(pRC)^{\frac{1}{2}}}\bigg[\sinh(pRC)^{\frac{1}{2}} -\sinh\left(\frac{1}{2}+\frac{x}{L}\right)(pRC)^{\frac{1}{2}} +\sinh\left(-\frac{1}{2}+\frac{x}{L}\right)(pRC)^{\frac{1}{2}}\bigg], \tag{20}$$

$$g(p)=(pRC)^{\frac{1}{2}}\sinh(pRC)^{\frac{1}{2}}, \tag{21}$$

$v=f(p)/g(p)$ satisfies the conditions for the use of the formula (18). The roots of the equation

$$g(p)=(pRC)^{\frac{1}{2}}\sinh(pRC)^{\frac{1}{2}}=0, \tag{22}$$

different from zero are

$$(p_nRC)^{\frac{1}{2}}=\pm jn\pi, \quad n=1, 2, 3, \cdots, \tag{23}$$

or

$$p_n=-\frac{n^2\pi^2}{RC}, \quad n=1, 2, 3, \cdots. \tag{24}$$

The time function $V=\mathfrak{L}^{-1}v$ is then

$$V(x, t)=S_0rL^2\Bigg\{\frac{1}{8}-\frac{1}{2}\left(\frac{x}{L}\right)^2 -\frac{4}{\pi^3}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^3}\cos\frac{x}{L}(2n+1)\pi \times\exp\left[-\frac{(2n+1)^2\pi^2t}{RC}\right]\Bigg\}. \tag{25}$$

For $t=0$ the term containing the sign of summation expresses the Fourier series of $\frac{1}{8}-\frac{1}{2}(x/L)^2$ in the interval $-L/2<x<L/2$. $V$ can thus be expressed in the following way

$$V(x, t)=S_0rL^2\frac{4}{\pi^3}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^3}\cos\frac{x}{L}(2n+1)\pi \times\left\{1-\exp\left[-\frac{(2n+1)^2\pi^2t}{RC}\right]\right\}. \tag{26}$$

For $x=0$ this leads to

$$V(0, t)=S_0rL^2\frac{4}{\pi^3}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^3} \times\left\{1-\exp\left[-\frac{(2n+1)^2\pi^2t}{RC}\right]\right\}. \tag{27}$$

$[V(0, t)/S_0rL^2]=V_1'$ has been plotted as a function of $t/RC$ in Fig. 3.

$I$ is calculated from (6) and (26)

$$I(x, t)=S_0L\frac{4}{\pi^2}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^2}\sin\frac{x}{L}(2n+1)\pi \times\left\{1-\exp\left[-\frac{(2n+1)^2\pi^2t}{RC}\right]\right\}. \tag{28}$$

For $x=\pm L/2$ this leads to

$$I\left(\pm\frac{L}{2}, t\right)=\pm S_0L\frac{4}{\pi^2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2} \times\left\{1-\exp\left[-\frac{(2n+1)^2\pi^2t}{RC}\right]\right\}. \tag{29}$$

$[I(\pm(L/2), t)/\pm S_0L]=I_1'$ has been plotted as a function of $t/RC$ in Fig. 4.

The amount of oil which flows through a cross section of the cable, $M$, is calculated by integrating the expression for $I$, given above.

$$M(x, t)=\int_0^t I(x, \lambda)d\lambda=S_0rcL^3\Bigg\{\frac{x}{L}\frac{t}{RC} -\frac{4}{\pi^4}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^4}\sin\frac{x}{L}(2n+1)\pi \times\left[1-\exp\left[-\frac{(2n+1)^2\pi^2t}{RC}\right]\right]\Bigg\}. \tag{30}$$

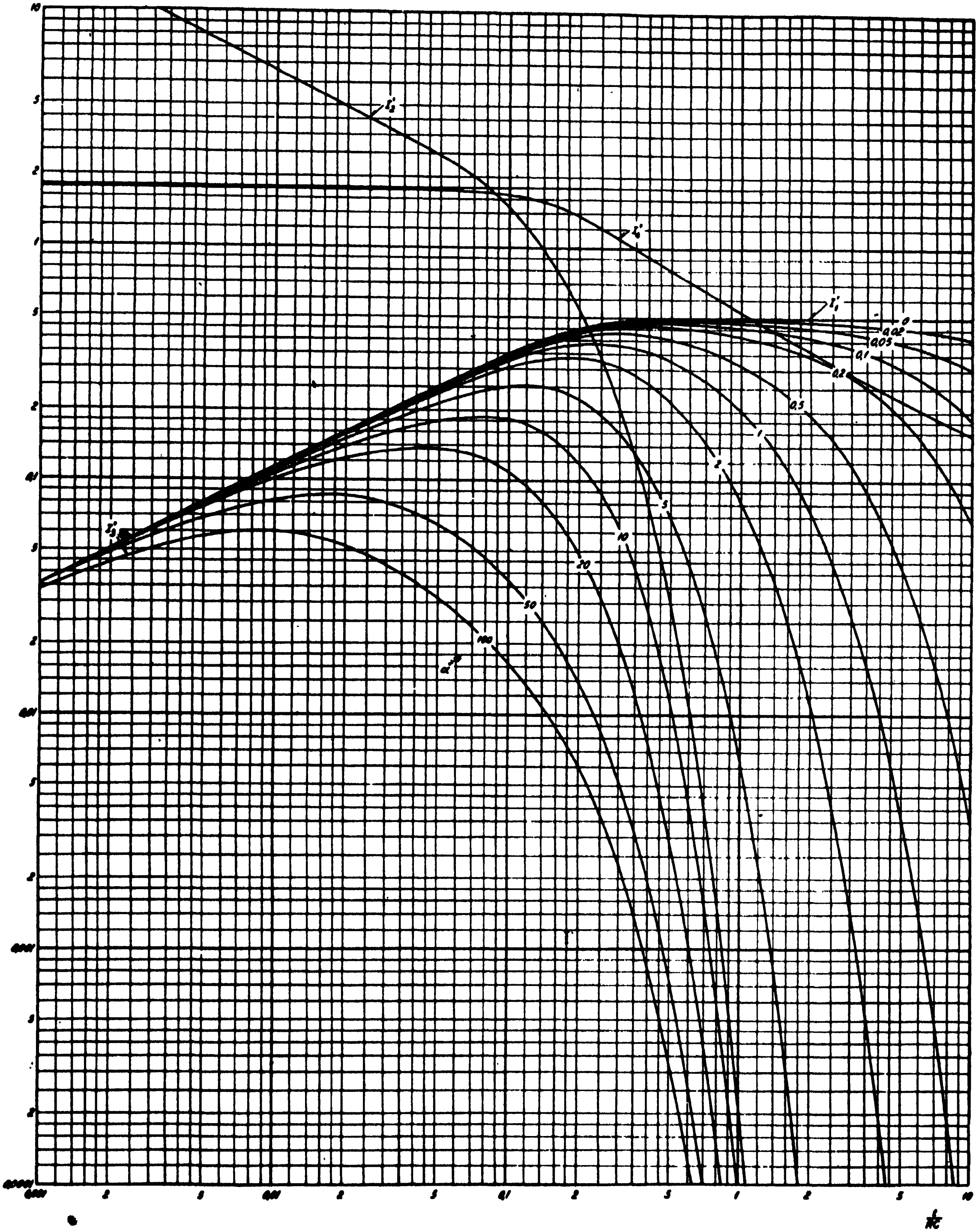


FIG. 4. Generalized functions used in the calculation of the oil current at the endpoints of a cable section with a time constant for the oil flow $RC$.

$I_1'(t/RC)$ applies in the case of an oil expansion $S(t)=1\cdot 1$.
$I_2'(t/RC)$ applies in the case of an oil expansion $S(t)=(d/dt)(1\cdot 1)$.
$I_3'(t/RC)$ applies in the case of an oil expansion $S(t)=e^{-\alpha t/RC}\cdot 1$.
$I_4'(t/RC)$ applies in the case of an oil expansion $S(t)=1/(t)^{\frac{1}{2}}\cdot 1$.

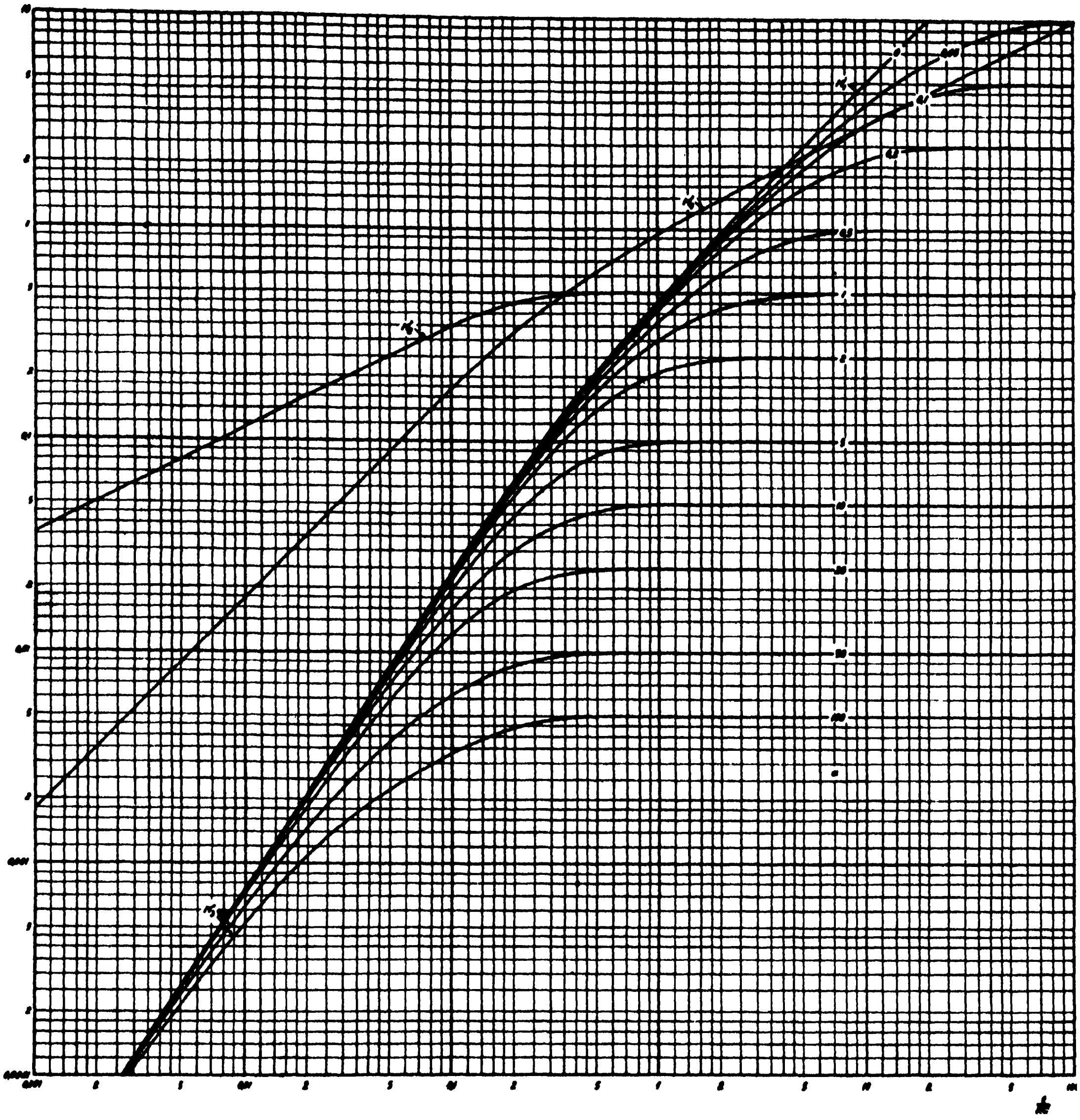


FIG. 5. Generalized functions used in the calculation of the amount of oil which passes the endpoints of a cable section with a time constant for the oil flow $RC$.

$M_1'(t/RC)$ applies in the case of an oil expansion $S(t)=1\cdot 1$.
$M_2'(t/RC)$ applies in the case of an oil expansion $S(t)=(d/dt)(1\cdot 1)$.
$M_3'(t/RC)$ applies in the case of an oil expansion $S(t)=e^{-\alpha t/RC}\cdot 1$.
$M_4'(t/RC)$ applies in the case of an oil expansion $S(t)=1/(t)^{\frac{1}{2}}\cdot 1$.

For $x=\pm L/2$ this leads to

$$M\left(\pm\frac{L}{2},t\right)=\pm S_0 rcL^3\left\{\frac{t}{2RC}-\frac{4}{\pi^4}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^4}\times\left[1-\exp\left[-\frac{(2n+1)^2\pi^2 t}{RC}\right]\right]\right\}. \quad (31)$$

$[M(\pm(L/2),t)/\pm S_0 rcL^3]=M_1'$ has been plotted as a function of $t/RC$ in Fig. 5.

In consequence of the linearity of the partial differential equations for the flow, the pressure, current and amount of oil which flows through a cross section of the cable for an arbitrary oil expansion $S(t)\cdot 1$ can be calculated from the corresponding solutions, given above, valid for an oil expansion of the form $S_0\cdot 1$. The linearity of the differential equations also causes the action (the pressure, the current, or the amount of oil) of an oil expansion given as a sum of time functions, to be obtained as the sum of the actions of the individual terms of which the oil expansion consists.

First it is necessary to mention some methods described in the literature for calculating the oil expansion in oil-filled cables for different types of load variations. Then the pressure, current, and amount of oil which flows through a cross section of the cable will be calculated for some important time functions of the oil expansion.

## OIL EXPANSION FOR DIFFERENT TYPES OF LOAD VARIATIONS

During the last fifty years a great many papers dealing with temperature variations in cables subjected to load variations have been published. For the purpose of calculating the oil expansion in cables subjected to ordinary load variations the methods suggested by Shanklin and Buller[10] and Miller and Wollaston[11] seem to be very useful. Using either of these methods the oil expansion is obtained as a sum of exponential terms $S_0e^{-t/T}\cdot 1$. From what has previously been mentioned, pressure, current, and amount of oil which flows through a cross section of the cable are calculated by summing the quantities corresponding to the individual exponential terms. In a later part of this paper, therefore, the pressure, current, and amount of oil for an oil expansion $S_0e^{-t/T}\cdot 1$ will be calculated.

The above mentioned methods for calculating the variations of temperature and oil expansion in oil-filled cables at load variations are rather inconvenient when the load variation in question is a short circuit. The temperature function in the case of short circuit is better computed on the basis of other simplifying assumptions, as the interesting part of the temperature variation in the case of a short circuit occurs in a relatively thin layer of the insulation near the conductor, in which the short-circuit current is flowing. Herlitz[12] has worked out a theory for temperature during and after a short circuit in conductors in contact with paper-oil insulation, whereby consideration has been given to the heat capacitance of the conductors and to the temperature coefficient of the conductor resistance. The oil-expansion computed on the basis of this theory is expressed by a function involving exponential functions and error integrals.

For use in the present paper we shall develop a very approximate and very simple expression for the oil expansion in an oil-filled cable in the case of short circuit in conductors in contact with paper-oil insulation. We hereby make the simplifying assumptions (1) that the conductors, in which the short-circuit current flows, have a plane surface, (2) that the insulation can be regarded as semi-infinite, (3) that the short circuit is instantaneous, and (4) that the temperature in the conductors remains constant after the short circuit has occurred. The surface of the conductors per unit length of the cable is called $f$, the specific heat-conduction resistance of the insulation, $p$, the specific heat of the insulation with respect to a unit volume, $q$, and the temperature, which the conductors attain at the short circuit, $\vartheta_0$. If $x$ denotes the distance from the surface of a conductor to a point of the insulation, the temperature field $\vartheta$ in the insulation is expressed by

$$\vartheta(x,t)=\vartheta_0\left[1-\phi\left(\frac{x(pq)^{1/2}}{2(t)^{1/2}}\right)\right]\cdot 1. \qquad (32)$$ [13]

The heat effect per unit length of the cable, $P$, transferred from the conductors to the insulation is then

$$P(t)=-\frac{f}{p}\left(\frac{\partial\vartheta}{\partial t}\right)_{x=0}=\frac{1}{(\pi)^{1/2}}f\vartheta_0\left(\frac{q}{p}\right)^{1/2}\frac{1}{(t)^{1/2}}\cdot 1. \qquad (33)$$

If we denote the effective volumetric expansion coefficient of the insulation by $\beta'$, the oil expansion per unit time per unit length of the cable, $S$, is

$$S(t)=\frac{\beta' P(t)}{q}=\frac{K_0}{(t)^{1/2}}\cdot 1 \qquad (34)$$

where

$$K_0=\frac{1}{(\pi)^{1/2}}\frac{\beta' f\vartheta_0}{(pq)^{1/2}}. \qquad (35)$$

---

[10] G. B. Shanklin and F. H. Buller, Gen. Elec. Rev. **34**, 523 (1931).

[11] K. W. Miller and F. O. Wollaston, Trans. A.I.E.E. **52**, 98 (1933).

[12] I. Herlitz, Teknisk Tidskrift, Elektroteknik **71**, 204 (1941).

[13] H. S. Carslaw, *Introduction to the Mathematical Theory of the Conduction of Heat in Solids* (Dover Publications, New York, 1945), second ed., p. 35.

Pressure, current, and amount of oil which flows through a cross section of the cable for an oil expansion according to the very approximate method of computation set forth here will be calculated in a following paragraph. The pressure so obtained will be larger than the actual pressure, especially in cables with small conductor areas, since the temperature of the conductor actually decreases after the short circuit instead of being constant as here supposed.

In a future paper, the pressure, current and amount of oil for the case of a short-circuit current flowing in conductors in contact with paper-oil insulation will be worked out for an oil expansion calculated by a more exact method of computation resembling that given by Herlitz.

The expansion of oil in canals, if any, in the conductors is difficult to calculate, since the heat transfer from the conductors to the oil in the canals takes place both by conduction and by convection. One is safe in assuming that the oil in a canal of any conductor has at any moment the same temperature as the conductor. In the case of an instantaneous short circuit the oil in a conductor canal will then be subjected to a sudden temperature rise resulting in a sudden expansion of the oil. If, as above, $\vartheta_0$ denotes the temperature rise of the conductors because of an instantaneous short circuit, and if $a$ denotes the volume of the conductor canal per unit length, and $\beta$ the volumetric expansion coefficient of the oil, the oil expansion per unit length per unit time, $S$, for the oil in the conductor canal in question is, using the terminology of the operational calculus, expressed as follows

$$S(t)=\frac{d}{dt}(N_0\cdot 1), \tag{36}$$

where

$$N_0=a\vartheta_0\beta. \tag{37}$$

Pressure, current, and amount of oil will be computed for an expansion of oil of this kind in the following paragraph. The quantities so obtained turn out to be more convenient for calculating pressure, current, and amount of oil for an arbitrary expansion of oil than the expressions obtained above for an oil expansion $S_0\cdot 1$.

## PRESSURE AND FLOW FOR AN INSTANTANEOUS OIL EXPANSION $(d/dt)(N_0\cdot 1)$

Pressure, current, and amount of oil are calculated by the expressions earlier obtained for an oil expansion $S_0\cdot 1$ by differentiating these with respect to time, and writing $N_0$ instead of $S_0$. In this way one obtains

$$V(x,t)=\frac{N_0}{c}\frac{4}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}\cos\frac{x}{L}(2n+1)\pi\exp\left[-\frac{(2n+1)^2\pi^2t}{RC}\right]. \tag{38}$$

For $x=0$ this leads to

$$V(0,t)=\frac{N_0}{c}\frac{4}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}\exp\left[-\frac{(2n+1)^2\pi^2t}{RC}\right]. \tag{39}$$

$[cV(0,t)/N_0]=V_2'$ has been plotted as a function of $t/RC$ in Fig. 3.

$$I(x,t)=\frac{N_0}{rcL}4\sum_{n=0}^{\infty}(-1)^n\sin\frac{x}{L}(2n+1)\pi\exp\left[-\frac{(2n+1)^2\pi^2t}{RC}\right]. \tag{40}$$

For $x=\pm L/2$ this leads to

$$I\left(\pm\frac{L}{2},t\right)=\pm\frac{N_0}{rcL}4\sum_{n=0}^{\infty}\exp\left[-\frac{(2n+1)^2\pi^2t}{RC}\right]. \tag{41}$$

$[rcLI(\pm(L/2), t)/\pm N_0] = I_2'$ has been plotted as a function of $t/RC$ in Fig. 4.

$$M(x, t) = N_0 L\left\{\frac{x}{L} - \frac{4}{\pi^2}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^2}\sin\frac{x}{L}(2n+1)\pi\exp\left[-\frac{(2n+1)^2\pi^2 t}{RC}\right]\right\}$$

$$= N_0 L\frac{4}{\pi^2}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^2}\sin\frac{x}{L}(2n+1)\pi\left\{1-\exp\left[-\frac{(2n+1)^2\pi^2 t}{RC}\right]\right\}. \tag{42}$$

For $x = \pm L/2$ this leads to

$$M\left(\pm\frac{L}{2}, t\right) = \pm N_0 L\frac{4}{\pi^2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}\left\{1-\exp\left[-\frac{(2n+1)^2\pi^2 t}{RC}\right]\right\}. \tag{43}$$

$[M(\pm(L/2), t)/\pm N_0 L] = M_2'$ has been plotted as a function of $t/RC$ in Fig. 5.

### PRESSURE AND FLOW FOR AN ARBITRARY OIL EXPANSION $S(t)\cdot 1$

If any action (pressure, current, or amount of oil) of an oil expansion $(d/dt)(1\cdot 1)$ is $A(t)$, then because of the linearity of the partial differential equations, valid for the oil flow, the action $G(t)$ of an oil expansion $S(t)$ is expressed by Duhamel's integral

$$G(t) = \int_0^t S(\lambda)A(t-\lambda)d\lambda. \tag{44}$$

Using this equation in connection with the above obtained expressions for pressure, current, and amount of oil for an oil expansion $(d/dt)(N\cdot 1)$, (38), (40), and (42), one obtains

$$V(x, t) = \frac{1}{c}\frac{4}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}\cos\frac{x}{L}(2n+1)\pi\exp\left[-\frac{(2n+1)^2\pi^2 t}{RC}\right]\int_0^t S(\lambda)\exp\left[\frac{(2n+1)^2\pi^2\lambda}{RC}\right]d\lambda, \tag{45}$$

$$I(x, t) = \frac{1}{rcL}4\sum_{n=0}^{\infty}(-1)^n\sin\frac{x}{L}(2n+1)\pi\exp\left[-\frac{(2n+1)^2\pi^2 t}{RC}\right]\int_0^t S(\lambda)\exp\left[\frac{(2n+1)^2\pi^2\lambda}{RC}\right]d\lambda, \tag{46}$$

$$M(x, t) = L\frac{4}{\pi^2}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^2}\sin\frac{x}{L}(2n+1)\pi$$

$$\times\left\{\int_0^t S(\lambda)d\lambda - \exp\left[-\frac{(2n+1)^2\pi^2 t}{RC}\right]\int_0^t S(\lambda)\exp\left[\frac{(2n+1)^2\pi^2\lambda}{RC}\right]d\lambda\right\}. \tag{47}$$

### PRESSURE AND FLOW FOR AN EXPONENTIALLY DECREASING OIL EXPANSION $S_0e^{-t/T}\cdot 1$

Pressure, current, and amount of oil for the oil expansion $S_0e^{-t/T}\cdot 1$ is calculated by substituting $S(t) = S_0e^{-t/T}$ in Eqs. (45), (46), and (47).

$$V(x, t) = S_0 rL^2\frac{4}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)((2n+1)^2\pi^2-(RC/T))}\cos\frac{x}{L}(2n+1)\pi$$

$$\times\left\{\exp\left(-\frac{RC}{T}\frac{t}{RC}\right)-\exp\left[-\frac{(2n+1)^2\pi^2 t}{RC}\right]\right\}. \tag{48}$$

For $x = 0$ this leads to

$$V(0, t) = S_0 rL^2\frac{4}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)((2n+1)^2\pi^2-(RC/T))}\left\{\exp\left(-\frac{RC}{T}\frac{t}{RC}\right)-\exp\left[-\frac{(2n+1)^2\pi^2 t}{RC}\right]\right\}. \tag{49}$$

$[V(0,t)/S_0rL^2]=V_3'$ has been plotted in Fig. 3 as a function of $t/RC$ for $\alpha=RC/T=0.02$, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, and 100.

$$I(x,t)=S_0L4\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^2\pi^2-(RC/T)}\sin\frac{x}{L}(2n+1)\pi\left\{\exp\left(-\frac{RC}{T}\frac{t}{RC}\right)-\exp\left[-\frac{(2n+1)^2\pi^2t}{RC}\right]\right\}. \quad (50)$$

For $x=\pm L/2$ this leads to

$$I\left(\pm\frac{L}{2},t\right)=\pm S_0L4\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2\pi^2-(RC/T)}\left\{\exp\left(-\frac{RC}{T}\frac{t}{RC}\right)-\exp\left[-\frac{(2n+1)^2\pi^2t}{RC}\right]\right\}. \quad (51)$$

$[I(\pm(L/2),t)/\pm S_0L]=I_3'$ has been plotted in Fig. 4 as a function of $t/RC$ for the same values of $\alpha=RC/T$ as given above.

$$M(x,t)=S_0rcL^3\left\{\frac{x}{L}\frac{T}{RC}\left[1-\exp\left(-\frac{RC}{T}\frac{t}{RC}\right)\right]-\frac{4}{\pi^2}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^2((2n+1)^2\pi^2-(RC/T))}\right.$$
$$\left.\times\sin\frac{x}{L}(2n+1)\pi\left[\exp\left(-\frac{RC}{T}\frac{t}{RC}\right)-\exp\left[-\frac{(2n+1)^2\pi^2t}{RC}\right]\right]\right\}. \quad (52)$$

For $x=\pm L/2$ this leads to

$$M\left(\pm\frac{L}{2},t\right)=\pm S_0rcL^3\left\{\frac{T}{2RC}\left[1-\exp\left(-\frac{RC}{T}\frac{t}{RC}\right)\right]-\frac{4}{\pi^2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2((2n+1)^2\pi^2-(RC/T))}\right.$$
$$\left.\times\left[\exp\left(-\frac{RC}{T}\frac{t}{RC}\right)-\exp\left[-\frac{(2n+1)^2\pi^2t}{RC}\right]\right]\right\}. \quad (53)$$

$[M(\pm(L/2),t)/\pm S_0rcL^3]=M_3'$ has been plotted in Fig. 5 as a function of $t/RC$ for the same values of $\alpha=RC/T$ as given earlier.

A practical example will demonstrate the use of the curves.

### Example 2

For a certain cable the oil expansion per unit length per unit time for a sudden switching on the load is

$$S(t)=S_1[e^{-(t/4000)}-0.6e^{-(t/400)}-0.2e^{-(t/100)}]$$

where $S_1$ is a constant. It is desired to calculate the variation of pressure at the midpoint of the cable, partly for switching on the load at the time $t=0$, after which the load remains switched on an infinite time, partly for switching on the load at the time $t=0$ and switching off the load at the time $t=200$ sec. The computation of the pressure will be carried out for a time constant of the cable $RC=0$, 1000, and 4000 sec.

For $RC=0$ the pressure $V$ in switching on the load is

$$(V(t))_{RC=0}=\tfrac{1}{8}S_1rL^2[e^{-(t/4000)}-0.6e^{-(t/400)}-0.2e^{-(t/100)}].$$

$[(V(t))_{RC=0}/\tfrac{1}{8}S_1rL^2]$ has been plotted in Fig. 6.

For the case where the load is switched off after 200 sec., the pressure $W$ is

$$(W(t))_{RC=0}=\begin{cases}(V(t))_{RC=0} & \text{for } t<200 \text{ sec.}\\ (V(t))_{RC=0}-(V(t-200))_{RC=0} & \\ & \text{for } t>200 \text{ sec.}\end{cases}$$

$[(W(t))_{RC=0}/\tfrac{1}{8}S_1rL^2]$ has been plotted in Fig. 7.

For $RC=1000$ sec. one obtains for the exponential terms in the expression for $S$

$$\frac{RC}{T_1}=\frac{1000}{4000}=0.25,$$

$$\frac{RC}{T_2}=\frac{1000}{400}=2.5,$$

$$\frac{RC}{T_3}=\frac{1000}{100}=10.$$

The pressure $V$ in switching on the load is, accordingly,

$$(V(t))_{RC=1000\ \text{sec.}}=S_1rL^2\left[\left(V_3'\left(\frac{t}{1000}\right)\right)_{\alpha=0.25}-0.6\left(V_3'\left(\frac{t}{1000}\right)\right)_{\alpha=2.5}-0.2\left(V_3'\left(\frac{t}{1000}\right)\right)_{\alpha=10}\right].$$

The functions $V_3'$ are taken from Fig. 3. $[(V(t))_{RC=1000\ \text{sec.}}/\frac{1}{8}S_1rL^2]$ has been plotted in Fig. 6.

The pressure $W$ for the load remaining switched on 200 sec. is calculated in the same way as in the case $RC=0$ treated above.

$[(W(t))_{RC=1000\ \text{sec.}}/\frac{1}{8}S_1rL^2]$ has been plotted in Fig. 7.

For $RC=4000$ sec. one obtains in a similar way

$$(V(t))_{RC=4000\ \text{sec.}}=S_1rL^2\left[\left(V_3'\left(\frac{t}{4000}\right)\right)_{\alpha=1}-0.6\left(V_3'\left(\frac{t}{4000}\right)\right)_{\alpha=10}-0.2\left(V_3'\left(\frac{t}{4000}\right)\right)_{\alpha=40}\right].$$

$[(V(t))_{RC=4000\ \text{sec.}}/\frac{1}{8}S_1rL^2]$ has been plotted in Fig. 6 and $[(W(t))_{RC=4000\ \text{sec.}}/\frac{1}{8}S_1rL^2]$ in Fig. 7.

Figure 6 shows that in the present case the variation in the computed pressure, resulting from taking into account the elasticity and the compressibility of the cable, is but little even with a time constant for the oil flow so large as 4000 sec. If all the terms in the expression for $S$ had the same sign, the variation in the calculated maximum pressure, however, would have been considerable, with the time constants for the oil flow occurring in the present case. Mostly, however, $S$ is composed of exponential terms with different signs in such a way that $S$ at the time $t=0$ assumes a small value, thereafter increases to a maximum, and then decreases to zero in resemblance to the case occurring in the example. With the values of the time constant for the oil flow, and the time constant in the first exponential term in the expression for the oil expansion occurring in practical cases, the maximum pressure in switching on a constant load calculated for $RC=0$ seldom diverges any

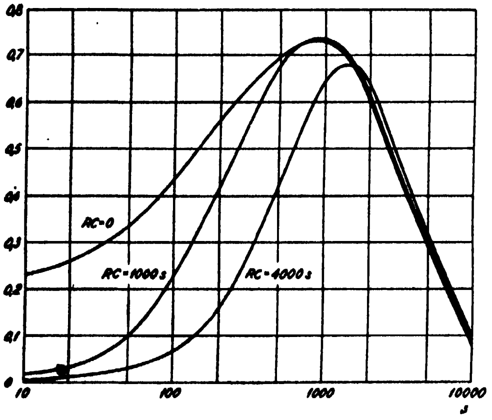


FIG. 6. Relative pressure at the midpoint of a cable in switching on a load giving an oil expansion $S(t)=S_1[e^{-t/4000}-0.6e^{-t/400}-0.2e^{-t/100}]$ for different time constants of the oil flow.

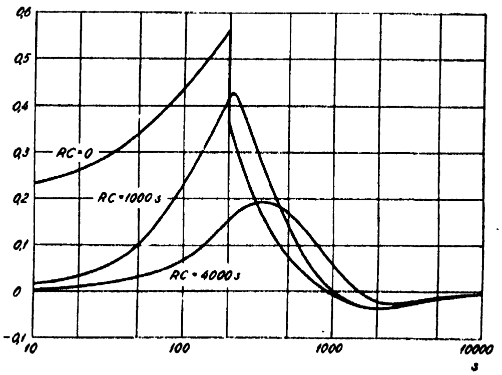


FIG. 7. Relative pressure at the midpoint of a cable in switching on for 200 sec. the same load as occurring in Fig. 6 for different time constants of the oil flow.

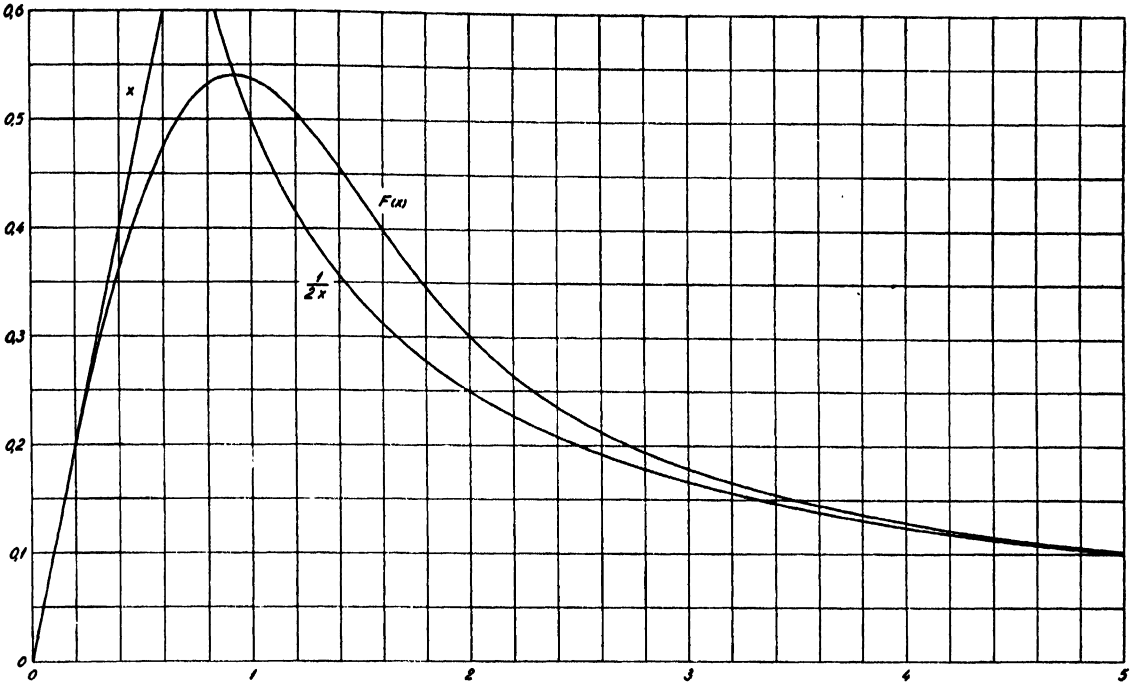


FIG. 8. $F(x)=e^{-x^2}\int_0^x e^{\xi^2}d\xi$ together with the approximations $F(x)\doteq x$ and $F(x)\doteq 1/2x$ valid for small, respectively for large values of $x$.

considerable amount from the pressure calculated for the actual value of $RC$.

This is not the case in the calculation of the maximum increase of pressure and the following maximum decrease of pressure occurring in switching on a constant load for a time interval that is small relative to the time constant for the oil flow in the cable. Figure 7 shows thus that the maximum pressure for the load in question with $RC=1000$ sec. comes to 76 percent, and with $RC=4000$ sec. only 34 percent of the maximum pressure with $RC=0$.

The following decrease of pressure is reduced too by the elasticity and the compressibility of the cable.

### PRESSURE AND FLOW FOR THE OIL EXPANSION $K_0/(t)^{\frac{1}{2}}\cdot 1$

Pressure, current, and amount of oil for the oil expansion $K_0/(t)^{\frac{1}{2}}\cdot 1$ is calculated by substituting $S(t)=K_0/(t)^{\frac{1}{2}}$ in Eqs. (45), (46), and (47).

The results in the present paragraph are most simply expressed by the function $F(x)$, defined in the following way:

$$F(x)=e^{-x^2}\int_0^x e^{\xi^2}d\xi. \tag{54}$$

This function has been tabulated by Miller and Gordon.[14] $F(x)$ has been plotted in Fig. 8 for $0<x<5$ together with the functions $y=2x$ and $y=1/2x$, which approximate the function $F(x)$ for small respectively large values of the argument $x$. The advantage in using the function $F(x)$ instead of the function

$$f(x)=\int_0^x e^{\xi^2}d\xi,$$

as used e.g., by Dawson,[15] is that $F(x)$ is limited and more convenient for interpolation than $f(x)$.[14]

Pressure, current, and amount of oil can now

---

[14] W. L. Miller and A. R. Gordon, J. Phys. Chem. **35**, 2785 (1931).

[15] H. G. Dawson; Proc. Lond. Math. Soc. **29**, 519 (1897–98).

be expressed in the following way:

$$V(x,t)=K_0\left(\frac{r}{c}\right)^{\frac{1}{2}}L\frac{8}{\pi^2}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^2}$$

$$\times\cos\frac{x}{L}(2n+1)\pi F\left((2n+1)\pi\left(\frac{t}{RC}\right)^{\frac{1}{2}}\right). \quad (55)$$

For $x=0$ this leads to

$$V(0,t)=K_0\left(\frac{r}{c}\right)^{\frac{1}{2}}L\frac{8}{\pi^2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}$$

$$\times F\left((2n+1)\pi\left(\frac{t}{RC}\right)^{\frac{1}{2}}\right). \quad (56)$$

$[V(0,t)/K_0(r/c)^{\frac{1}{2}}L]=V_4'$ has been plotted as a function of $t/RC$ in Fig. 3.

$$I(x,t)=\frac{K_0}{(rc)^{\frac{1}{2}}}\frac{8}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}$$

$$\times\sin\frac{x}{L}(2n+1)\pi F\left((2n+1)\pi\left(\frac{t}{RC}\right)^{\frac{1}{2}}\right). \quad (57)$$

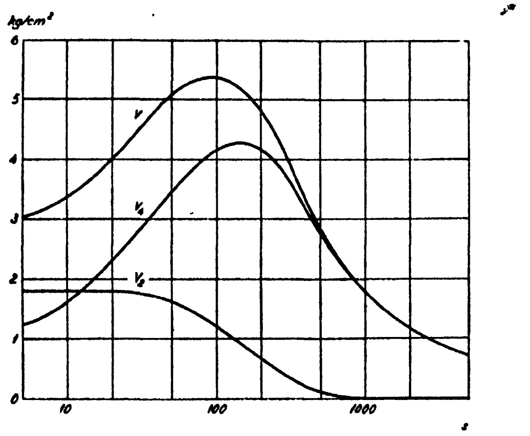


FIG. 9. Pressure $V$ at the midpoint of a cable, occurring at a short circuit in the cable as specified in the text, for 20°C temperature of the oil in the oil canal. $V=V_2+V_4$, where $V_2$ corresponds to the expansion of the oil in the canal, $V_4$ to the expansion of the oil in the insulation.

For $x=\pm L/2$ this leads to

$$I\left(\pm\frac{L}{2},t\right)=\pm\frac{K_0}{(rc)^{\frac{1}{2}}}\frac{8}{\pi}\sum_{n=0}^{\infty}\frac{1}{2n+1}$$

$$\times F\left((2n+1)\pi\left(\frac{t}{RC}\right)^{\frac{1}{2}}\right). \quad (58)$$

$[I(\pm(L/2),t)/\pm(K_0/(rc)^{\frac{1}{2}})]=I_4'$ has been plotted as a function of $t/RC$ in Fig. 4.

$$M(x,t)=K_0(rc)^{\frac{1}{2}}L^2\left[2\frac{x}{L}\left(\frac{t}{RC}\right)^{\frac{1}{2}}-\frac{8}{\pi^3}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^3}\right.$$

$$\left.\times\sin\frac{x}{L}(2n+1)\pi F\left((2n+1)\pi\left(\frac{t}{RC}\right)^{\frac{1}{2}}\right)\right]. \quad (59)$$

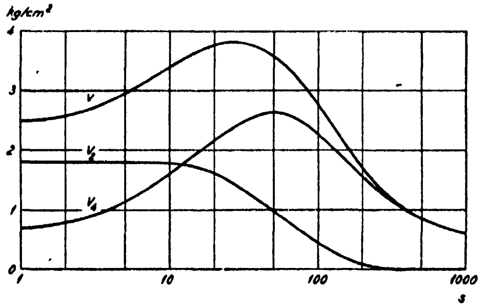


FIG. 10. Pressure $V$ at the midpoint of the same cable and for the same short circuit current in this cable as in Fig. 9 for 40°C temperature of the oil in the oil canal. $V=V_2+V_4$, where $V_2$ corresponds to the expansion of the oil in the canal, $V_4$ to the expansion of the oil in the insulation.

For $x=\pm L/2$ this leads to

$$M\left(\pm\frac{L}{2},t\right)=\pm K_0(rc)^{\frac{1}{2}}L^2\left[\left(\frac{t}{RC}\right)^{\frac{1}{2}}\right.$$

$$\left.-\frac{8}{\pi^3}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^3}F\left((2n+1)\pi\left(\frac{t}{RC}\right)^{\frac{1}{2}}\right)\right]. \quad (60)$$

$[M(\pm(L/2),t)/\pm K_0(rc)^{\frac{1}{2}}L]=M_4'$ has been plotted as a function of $t/RC$ in Fig. 5.

The use of the formulas worked out in this paragraph is explained by the following example.

### Example 3

It is desired to calculate the maximal pressure rise in the single-core cable described in Ex. 1 for a cable length of 1 km in the case of a short circuit of 1 sec. duration, during which the conductor and the oil in the conductor canal are heated 10°C. It is supposed that the heat effect of the secondary current in the sheath is negligible. The pressure will be calculated for the temperature of the oil in the canal 20°C and 40°C.

Besides the data for the cable given in Ex. 1

the following data are needed:

$$p=550°\text{C cm } W^{-1},\quad q=2°\text{C}^{-1}\text{ cm}^{-3}\ W\text{ sec.},$$
$$\beta=0.0007°\text{C}^{-1},\quad \beta'=0.0004°\text{C}^{-1}.$$

From the data given in Ex. 1 are further calculated

$$f=\pi\cdot 1.3=4.1\text{ cm},$$
$$a=(\pi/4)\cdot 0.6^2=0.28\text{ cm}^2.$$

The increase in pressure at the midpoint of the cable $V_2(0, t)$ caused by the expansion of the oil in the canal is according to (37) and (39)

$$V_2(0, t)=a\vartheta_0\beta\frac{1}{c}V_2'\left(\frac{t}{RC}\right)$$
$$=0.28\cdot 10\cdot 0.0007\cdot\frac{1}{1.09\cdot 10^{-3}}V_2'\left(\frac{t}{RC}\right)$$
$$=1.8V_2'\left(\frac{t}{RC}\right)\text{ kg cm}^{-2}$$

where $V_2'(t/RC)$ is taken from Fig. 3.

The increase in pressure at the midpoint of the cable $V_4(0, t)$ caused by the heat transferred from the conductor to the insulation is according to (35) and (56)

$$V_4(0, t)=\frac{\beta' f\vartheta_0 L(r)^{\frac{1}{2}}}{(\pi pqc)^{\frac{1}{2}}}V_4'\left(\frac{t}{RC}\right)$$
$$=\frac{0.0004\cdot 4.1\cdot 10\cdot 10^5 V_4'(t/RC)}{(\pi\cdot 550\cdot 2\cdot 1.09\cdot 10^{-3})^{\frac{1}{2}}}\begin{cases}(1.43\cdot 10^{-4})^{\frac{1}{2}}\\(5.3\cdot 10^{-4})^{\frac{1}{2}}\end{cases}$$
$$=10.1V_4'\left(\frac{t}{RC}\right)\text{ at } 20°\text{C}$$
$$=6.2V_4'\left(\frac{t}{RC}\right)\text{ at } 40°\text{C}$$

where $V_4'(t/RC)$ is taken from Fig. 3.

$V_2$, $V_4$, and $V=V_2+V_4$ for 20°C temperature of the oil in the oil canal have been plotted in Fig. 9, and $V_2$, $V_4$, and $V=V_2+V_4$ for 40°C oil temperature in Fig. 10.

From Fig. 9 and Fig. 10 is read the maximum pressure

$$V_{\max}=\begin{cases}5.4\text{ kg cm}^{-2}\text{ for } t=95\text{ sec. with } 20°\text{C}\\3.8\text{ kg cm}^{-2}\text{ for } t=30\text{ sec. with } 40°\text{C},\end{cases}$$

In working out the theory for the oil expansion after a short circuit we made the simplifying assumption that the temperature of the conductors remained constant after the short circuit. It is interesting to make an estimate in the present case of how much the conductor temperature has decreased from the moment of short circuit to the moment of occurrence of maximum pressure.

With the idealizing assumption mentioned above, namely, that the conductor temperature remains constant, the effect per unit length of the cable, $P$, transferred to the insulation was given in (33). The amount of heat per unit length, $Q$, transferred to the insulation from the moment of short circuit to the time $t$ is then

$$Q(t)=\int_0^t P(\lambda)d\lambda=\frac{2}{(\pi)^{\frac{1}{2}}}f\vartheta_0\left(\frac{q}{p}\right)^{\frac{1}{2}}(t)^{\frac{1}{2}}.$$

Hereby we calculate the amount of heat per unit length in the present case transferred to the insulation from the moment of short circuit to the moment $t_1$, when the maximum pressure occurs

$$Q(t_1)=\frac{2}{(\pi)^{\frac{1}{2}}}\cdot 4.1\cdot 10\cdot\left(\frac{2}{550}\right)^{\frac{1}{2}}\cdot\begin{cases}(95)^{\frac{1}{2}}=26\text{ W sec. cm}^{-1}\text{ for } 20°\text{C}.\\(30)^{\frac{1}{2}}=15\text{ W sec. cm}^{-1}\text{ for } 40°\text{C}.\end{cases}$$

These transferred amounts of heat have to be compared with the amount of heat per unit length, $Q_1$, accumulated in the conductor and in the oil canal. The specific heat for copper in respect to a unit volume is 3.4°C$^{-1}$ cm$^{-3}$ W sec. Neglecting the amount of heat accumulated in the oil canal we get

$$Q_1=0.95\cdot 3.4\cdot 10=32\text{ W sec. cm}^{-1}.$$

The amount of heat transferred from the conductor to the insulation from the moment of short circuit to the moment, when the maximal pressure, calculated according to the simplifying assumption mentioned above, occurs, is consequently not negligible compared with the total amount of heat accumulated in the conductor. The preparation of a more exact theory, as mentioned above, seems to be justified.

The work here presented has been carried out at Sieverts Kabelverk, Sundbyberg, Sweden. I thank Sieverts Kabelverk and especially Mr. H. Spanne, chief engineer at Sieverts Kabelverk, for permission to publish this paper.

# The Electrostatic Field of a Point Charge Inside a Cylinder, in Connection with Wave Guide Theory

C. J. Bouwkamp and N. G. de Bruijn
*Natuurkundig Laboratorium der N. V. Philips' Gloeilampenfabrieken, Eindhoven, Netherlands*

(Received February 21, 1947)

In this paper the field of a point charge inside a hollow, infinitely long circular cylinder is studied. The case of an axial point charge is treated in detail. Three different methods are developed. The first method shows how to calculate the induced charges at the surface of the cylinder without explicit knowledge of the potential itself. The surface charge-density function is obtained as the solution of a Fourier-type integral equation. Then the potential caused by these charges is calculated. The second method works in the opposite direction. Here the potential is obtained as solution of a boundary value problem, followed by the calculation of the corresponding charges at the surface of the cylinder.

The integral, obtained for the surface charge density, is transformed by contour integration. Although the resulting series is very useful for numerical purposes, a stronger method is necessary, in order to calculate the charge density just opposite the point source. Figure 1 shows the calculated values of the induced surface charge density. Several approximations for the charge-density function are considered, in connection with recent work of Weber (cf. Table I).

The same is done for the field inside the cylinder. Various formulas are given which allow of numerical calculations. In Table II some calculated values of the potential are shown.

The third method is based on the theory of Fourier-Bessel-Dini series. The potential is developed in terms of discrete normal solutions of the potential equation in cylindrical coordinates. The coefficients in this development can be derived from the behavior of the potential in the immediate neighborhood of the primary source.

Furthermore, it is emphasized that the study of the above potential problem can serve as a guide in questions of wave propagation in hollow circular cylinders. In this connection the third method is shown to be extremely useful, as it enables us to calculate directly the fields in the far zone from that in the immediate neighborhood of the exciting source. This new method is demonstrated in case of acoustic waves inside a cylinder, caused by a harmonically vibrating point source.

## 1. INTRODUCTION

PREVIOUS works of Schelkunoff,[1] Weyrich,[2] Buchholz,[3] and Oberhettinger,[4] among others, already contain complete information with regard to fields in circular pipes, as produced by Hertzian dipoles. Therefore, the generation of special types of wave patterns, caused by given current and charge distributions on the exiting antennas, can be dealt with quantitatively. The calculations in this domain of electromagnetics, however, are far from elementary, such that a non-mathematically minded engineer might hardly recognize the usefulness of those theoretical treatments. Hence, every attempt to diminish the necessary amount of mathematics, may be worth while to consider. Until now a considerable amount of complex function theory is required. We believe, however, that this can be greatly avoided if one bases oneself on the theory of Fourier-Bessel-Dini series. This method is undoubtedly within reach of every engineer,

---

[1] S. A. Schelkunoff, "Modified Sommerfeld's integral and its applications," Proc. I.R.E. **24**, 1388 (1936); and "Transmission theory of plane electromagnetic waves," *ibid.* **25**, 1457 (1937); with bibliography.

[2] R. Weyrich, "Ueber einige Randwertprobleme, insbesondere der Elektrodynamik," J. reine angew. Math. **172**, 133 (1934).

[3] H. Buchholz, "Gekoppelte Strahlungsfelder im kreiszylindrischen Hohlleiter," Ann. d. Physik **39**, 81 (1941).

[4] F. Oberhettinger, "Ueber ein Randwertproblem der Wellengleichung in Zylinderkoordinaten," Ann. d. Physik **43**, 136 (1943).

because the analogous conventional theory of the ordinary Fourier series is common knowledge nowadays.

In this paper we will not consider the vector wave propagation as occurring in electromagnetic theory. Those topics will be treated elsewhere. The new method will be demonstrated in the case of scalar wave propagation only.

The major part of the present article, however, will be devoted to the study of the electrostatic field of a point charge inside the cylinder. Consequently, only a minor part of the mathematics developed here is claimed to be useful in wave guide theory.

Beginning with the simplest problem, we confine ourselves to the case of a point charge situated on the axis of the cylinder. Later on we will briefly indicate how to extend the theory to excentric point charges.

## 2. STATEMENT OF THE SIMPLEST PROBLEM. CALCULATION OF THE CHARGES INDUCED ON THE SURFACE OF THE CYLINDER

Let $a$ be the radius of the cylinder, and let the point charge have a value $+1$. Furthermore, let $z$, $\rho$, $\varphi$ denote the right-handed cylindrical coordinates of a field point $P$. Henceforth the origin may coincide with the given point charge, and the $z$-axis be taken along the axis of the cylinder.

The electric force $\mathbf{E}$ can be derived from a potential $V$ according to $\mathbf{E}=-\operatorname{grad} V$. It is evident that the $z$-axis is an axis of symmetry, so that $V$ depends only on $z$ and $\rho$. In addition, $V$ is even in $z$.

To calculate $\mathbf{E}$ and $V$, we first follow a rather unconventional method, by calculation of the *charges*, induced on the wall of the cylinder, without explicit knowledge of the potential itself. Afterwards the field is readily found from the original point charge and these induced charges. The surface density of the induced charges will be denoted by $\eta(z)$.

We may remark that $V$ is the so-called Green function of the first kind. It is a regular solution of the potential equation, except for a singularity in the origin, so that the function $V(P)-1/r$ remains finite for $r\rightarrow 0$; furthermore, $V=0$ on the boundary, and both $V$ and $\operatorname{grad} V$ tend to zero, if $P$ tends to infinity inside the cylinder.

As is well known, there is one, and only one, function $V(z, \rho)$ fulfilling all these conditions simultaneously. Once this function is obtained, the charges on the boundary of the cylinder can be computed from

$$\eta(z)=\frac{1}{4\pi}\frac{\partial V}{\partial \rho}, \quad (\rho=a). \tag{1}$$

Let us now show how $\eta(z)$ can be found without explicit knowledge of the potential $V$. Be $g(z, \rho)$ a given regular solution of the potential equation inside the cylinder, and suppose $g$ and $\partial g/\partial z$ to be bounded in that domain. By Green's theorem, we at once express the value of $g$ in the origin as a surface integral over its boundary values, namely

$$g(0, 0)=-\frac{1}{4\pi}\int g\frac{\partial V}{\partial \rho}d\sigma=-\int g(z, a)\eta(z)d\sigma. \tag{2}$$

In this $d\sigma$ means the surface element of the boundary of the cylinder; the integration in (2) extends over this boundary.

Omitting the left-hand integral in (2), the remaining equality can be interpreted as an integral equation for $\eta(z)$, which may be readily solved, if a convenient choice is made for the rather arbitrary function $g$. We take

$$g(z, \rho)=e^{-i\lambda z}I_0(\lambda\rho) \tag{3}$$

where $I_0$ is the Bessel function of purely imaginary argument and zeroth order. The function (3) is a regular potential, with $g$ and $\partial g/\partial z$ bounded inside the cylinder, if the constant $\lambda$ is restricted to *real* values.

Substitution of (3) in (2) leads to

$$1=-\int_0^{2\pi} a d\varphi\int_{-\infty}^{\infty} e^{-i\lambda z}I_0(\lambda a)\eta(z)dz,$$

whence it follows

$$\int_{-\infty}^{\infty} z^{-i\lambda z}\eta(z)dz=-\frac{1}{2\pi a I_0(\lambda a)}, \quad (-\infty<\lambda<\infty). \tag{4}$$

Thus the integral equation (2) is reduced to one of the common Fourier type (4). Its solution can be written down immediately. By inversion of

(4), one obtains for the required charge density

$$\eta(z) = -\frac{1}{4\pi^2 a}\int_{-\infty}^{\infty} e^{i\lambda z}\frac{d\lambda}{I_0(\lambda a)}$$

$$= -\frac{1}{2\pi^2 a^2}\int_0^{\infty}\frac{\cos(z/a)t}{I_0(t)}dt. \qquad (5)$$

Starting from this integral expression, we can develop solutions more useful for numerical purposes. For non-vanishing $z$ we may transform the above integral into a series by means of the conventional contour integration. In the cut plane $z=0$, however, stronger methods must be used, as will be seen below.

Consider the following integral in the complex $t$-plane:

$$\frac{1}{2\pi i}\int_W \frac{e^{izt/a}}{I_0(t)}dt,$$

the path of integration $W$ leading from $-R$ to $R$ along the real axis, and back from $R$ to $-R$ along a semi-circle of radius $R$ and center 0. Avoiding values of $R$ for which the semi-circle passes through a zero of $I_0(t)$, we can prove that for positive values of $z$, the contribution of the semi-circle tends to zero, when $R$ tends to infinity. Let $k_1$, $k_2$, $\cdots$ denote, in ascending order, the positive zeros of the Bessel function $J_0(k)$. Then $ik_n$ are the simple poles of the integrand in the upper half of the $t$-plane. The residue theorem now gives

$$\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{e^{izt/a}}{I_0(t)}dt = \sum_{n=1}^{\infty}\frac{e^{-k_n z/a}}{I_0'(ik_n)}.$$

We have $I_0'(ik_n) = iJ_1(k_n)$, and so for the charge density

$$\eta(z) = -\frac{1}{2\pi a^2}\sum_{n=1}^{\infty}\frac{e^{-k_n z/a}}{J_1(k_n)}, \quad (z \geqslant 0). \qquad (6)$$

Only for $z=0$ does this series not converge in the ordinary sense. Unless $z$ is very small, it is very useful for numerical computation. A small number of terms are already sufficient in the range $z \geqslant a$.

The total charge, induced on the cylinder between $z \geqslant 0$ and $\infty$, can be found by integration:

$$2\pi a\int_z^{\infty}\eta(\zeta)d\zeta = -\sum_{n=1}^{\infty}\frac{e^{-k_n z/a}}{k_n J_1(k_n)}, \quad (z \geqslant 0).$$

The latter series does converge at $z=0$. The total amount of induced charge on the upper half of the cylinder should be $-\frac{1}{2}$; we have thus an opportunity to verify a well-known formula in Bessel-function theory, originally credited to Nielsen:

$$\sum_{n=1}^{\infty}\frac{1}{k_n J_1(k_n)} = \tfrac{1}{2}.$$

For small values of $z$ the convergence of the series (6) is too slow for practical evaluation. Hence we have to develop other expressions for $\eta(z)$, which do allow of numerical computation near the "cut plane" $z=0$. This can be accomplished by Watson's method for an analogous problem.[5]

Therefore we consider the following integral

$$\frac{1}{2\pi i}\int\frac{\cos\lambda w}{(w-t)I_0(w)\cos(\pi w/\beta)}dw$$

wherein $\lambda$, $\beta$, $t$ are real parameters; $\beta>0$, $\cos(\pi t/\beta)\neq 0$. The path of integration is a circle of radius $R>t$, and center 0, not passing through any zero of the denominator of the integrand. When one adds the condition $\beta|\lambda|<\pi$, the integral tends to zero, if $R$ increases to infinite values. Thus the sum of the residues of the integrand is zero also. The poles of the integrand are all simple; they lie at

$$w=t, \qquad \text{(a)}$$

$$w=\pm ik_n, \quad (n=1, 2, \cdots), \qquad \text{(b)}$$

$$w=\pm(\nu+\tfrac{1}{2})\beta, \quad (\nu=0, 1, 2, \cdots). \qquad \text{(c)}$$

The corresponding residues are easily found to be

$$\frac{\cos\lambda t}{I_0(t)\cos(\pi t/\beta)}, \qquad \text{(a}')$$

$$-\frac{\cosh\lambda k_n}{J_1(k_n)\cosh(\pi k_n/\beta)[k_n \pm it]}, \qquad \text{(b}')$$

$$-\frac{\beta}{\pi}\frac{(-1)^\nu\cos(\lambda\beta\nu+\lambda\beta/2)}{I_0(\beta\nu+\beta/2)[(\nu+\frac{1}{2})\beta\mp t]}. \qquad \text{(c}')$$

---

[5] G. N. Watson, "The use of series of Bessel functions in problems connected with cylindrical wind-tunnels," Proc. Roy. Soc. **130**, 29 (1930/31).

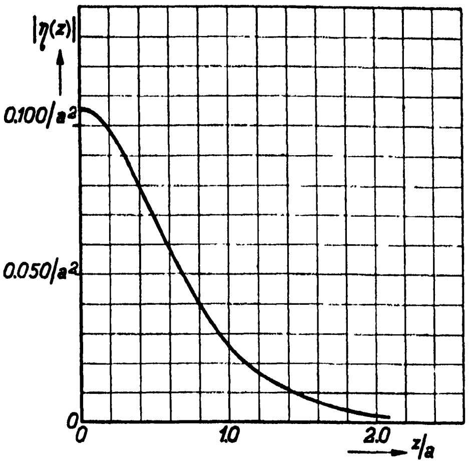


FIG. 1. Curve showing the surface charge density $|\eta|$ on the wall of the grounded cylinder, induced by a unit-point charge situated in the axis at $z=0$; $z$ denotes the distance along the cylinder, measured from the source containing plane; $a$ is the radius of the cylinder.

The sum of all residues together being zero, one has

$$\frac{\cos\lambda t}{I_0(t)\cos(\pi t/\beta)}=\sum_{n=1}^{\infty}\frac{2k_n\cosh\lambda k_n}{J_1(k_n)\cosh(\pi k_n/\beta)}\frac{1}{k_n{}^2+t^2}+\frac{\beta}{\pi}\sum_{\nu=0}^{\infty}\frac{(-1)^\nu\cos(\lambda\beta\nu+\lambda\beta/2)}{I_0(\beta\nu+\beta/2)}\frac{(2\nu+1)\beta}{(\beta\nu+\beta/2)^2-t^2}. \quad (7)$$

Evidently, the right-hand side of (7) is nothing else but the expansion in rational fractions of the left-hand side.

We now multiply both sides of (7) by $\cos(\pi t/\beta)$, thus neutralizing the poles in (c), and integrating from $t=-\infty$ to $t=\infty$. Using the following formulas[5]

$$\int_{-\infty}^{\infty}\frac{\cos(\pi t/\beta)}{k_n{}^2+t^2}dt=\frac{\pi}{k_n}e^{-\pi k_n/\beta},$$
$$\int_{-\infty}^{\infty}\frac{\cos(\pi t/\beta)}{(\beta\nu+\beta/2)^2-t^2}dt=\frac{(-1)^\nu 2\pi}{(2\nu+1)\beta}, \quad (8)$$

we easily obtain

$$\int_0^{\infty}\frac{\cos\lambda t}{I_0(t)}dt=\beta\sum_{\nu=0}^{\infty}\frac{\cos\lambda\beta(\nu+\frac{1}{2})}{I_0(\beta\nu+\beta/2)}+2\pi\sum_{n=1}^{\infty}\frac{\cosh\lambda k_n}{J_1(k_n)[1+e^{2\pi k_n/\beta}]}. \quad (9)$$

Hence, for the induced charge density

$$-a^2\eta(z)=\frac{\beta}{2\pi^2}\sum_{\nu=0}^{\infty}\frac{\cos(\beta\nu+\beta/2)z/a}{I_0(\beta\nu+\beta/2)}+\frac{1}{\pi}\sum_{n=1}^{\infty}\frac{\cosh k_n z/a}{J_1(k_n)[1+e^{2\pi k_n/\beta}]}. \quad (10)$$

It may be emphasized that our proof of (10) holds for $\beta>0$ and $-\pi a/\beta<z<\pi a/\beta$. The first series on the right of (10) is convergent for all real values of $z$. It may be considered as an approximation of the integral in (5), obtained when the integration interval is divided into equal parts of length $\beta$, and the integrand in each separate interval is replaced by its central value. The second series should then be interpreted as a correction term. The latter, however, only converges for $-2\pi a/\beta<z<2\pi a/\beta$. By analytic continuation, it can be shown that (10) holds for this extended range, twice as large as might be expected from the proof of (10), as given above. Moreover, for the special value $\beta=a\pi/|z|$, the first series vanishes identically, whilst the "correction term" reduces to the series (6). Consequently, Eq. (10) is more basic than (6).

Equation (10) is very suitable for numerical calculations. The constant $\beta$ is still arbitrary, except for $-2\pi a/\beta<z<2\pi a/\beta$. Let us consider the extreme case $z=0$ in detail. We already saw that in this case (6) cannot be used on behalf of its divergence. Equation (10) can be used with any positive $\beta$. We calculated $-2\pi^2a^2\eta(0)$ in seven decimals with $\beta=1$, $\beta=3$, respectively. In the first case we had to take eighteen terms of the first series and only one of the second. The result was found to be

$$2.0832296+0.0000033=2.0832329.$$

With $\beta=3$ we had to consider six terms of the first, and three terms of the second series, obtaining in this case

$$2.0053020+0.0779307=2.0832327,$$

in close agreement with the earlier value. Thus we can state

$$\int_0^{\infty}\frac{dt}{I_0(t)}=2.083232\cdots,\quad \eta(0)=-\frac{0.10554}{a^2}. \quad (11)$$

Numerical computation of the surface charge density is now easily established. When $z \geqslant a$, already the first four terms of (6) are sufficient to guarantee the correctness of the fourth decimal. Inserting known data from existing Bessel-function tables, (6) becomes

$$\begin{aligned}-a^2\eta(z) = 0.30657 \exp(-2.40z/a)\\ -0.46774 \exp(-5.52z/a)\\ +0.58636 \exp(-8.65z/a)\\ -0.68466 \exp(-11.79z/a). \quad (12)\end{aligned}$$

In the remaining range $0 \leqslant z \leqslant a$, Eq. (10) may be used with $\beta = 2$. Then one obtains

$$\begin{aligned}-a^2\eta(z) = 0.08003 \cos(z/a) + 0.02076 \cos(3z/a)\\ +0.00372 \cos(5z/a) + 0.00060 \cos(7z/a)\\ +0.00009 \cos(9z/a) + 0.00001_4 \cos(11z/a)\\ +0.00032 \cosh(2.4z/a). \quad (13)\end{aligned}$$

The data of Fig. 1 were computed by means of Eqs. (12) and (13). The curve shows a rapid decrement of the surface charges with increasing distance from the cut plane.

By expanding both sides of (9) into a power series of $\lambda$, and identifying the corresponding coefficients, we obviously obtain

$$\int_0^\infty \frac{t^{2r}}{I_0(t)}dt = \beta \sum_{\nu=0}^{\infty} \frac{(\beta\nu+\beta/2)^{2r}}{I_0(\beta\nu+\beta/2)} + (-1)^r 2\pi \sum_{n=1}^{\infty} \frac{k_n^{2r}}{J_1(k_n)[1+e^{2\pi k_n/\beta}]}.$$

The numerical values of these integrals can be found in the same way as shown above for $r = 0$. Then for $\eta(z)$ a power series in $z^2/a^2$ can be derived, convergent for $|z| < a$. The expression (13), however, is much more useful for numerical purposes in the corresponding range of $z$-values.

## 3. FORMULAS FOR THE POTENTIAL AND THE ELECTRIC FIELD

The potential $V(P)$ can be considered as the sum of the potential $1/r$, due to the unit charge in the origin, and the potential $V_0(P)$, due to the induced charges on the surface of the cylinder:

$$V(P) = \frac{1}{r} + V_0(P) = \frac{1}{} + \int \frac{\eta(\zeta)}{} d\sigma.$$

In this $r$, $r_1$, denote the distances from $P$ to the origin, and the surface element, respectively. Again, $V_0(P)$ depends only upon $z$ and $\rho$.

At first we restrict ourselves to points on the axis of the cylinder. Then we have

$$V_0(z, 0) = 2\pi a \int_{-\infty}^{\infty} \frac{\eta(x+u)}{(u^2+a^2)^{\frac{1}{2}}} du.$$

Using (5), we obtain

$$V_0(z, 0) = -\frac{2}{\pi} \int_0^\infty \frac{\cos\lambda z}{I_0(\lambda a)} d\lambda \int_0^\infty \frac{\cos\lambda u}{(u^2+a^2)^{\frac{1}{2}}} du.$$

On account of a well-known integral for the $K$-function, namely

$$K_0(\lambda a) = \int_0^\infty \frac{\cos\lambda u}{(u^2+a^2)^{\frac{1}{2}}} du, \quad (14)$$

we obtain for the potential of the induced charges, at points on the axis of the cylinder

$$V_0(z, 0) = -\frac{2}{\pi} \int_0^\infty \frac{K_0(\lambda a)}{I_0(\lambda a)} \cos\lambda z d\lambda. \quad (15)$$

In the rotationally symmetrical problem under consideration, the potential outside the axis is easily obtained, once the potential on the axis of symmetry has been found. We simply have to take the mean value of $V_0(z+i\rho\cos\varphi, 0)$ over a full period of the angle $\varphi$. Thus

$$V_0(z, \rho) = \frac{1}{2\pi} \int_0^{2\pi} V_0(z+i\rho\cos\varphi, 0) d\varphi. \quad (16)$$

The integration over $\varphi$ can be performed, leading to

$$V_0(z, \rho) = -\frac{2}{\pi} \int_0^\infty \frac{K_0(\lambda a)}{I_0(\lambda a)} I_0(\lambda\rho) \cos\lambda z d\lambda, \quad (17)$$

and for the total potential we get

$$V(z, \rho) = \frac{1}{(z^2+\rho^2)^{\frac{1}{2}}} - \frac{2}{\pi a} \int_0^\infty \frac{K_0(t)}{I_0(t)} \times I_0(t\rho/a) \cos(tz/a) dt. \quad (18)$$

The corresponding axial and radial components of the electric force are readily derived from

(18) by differentiation

$$E_z=-\frac{z}{(z^2+\rho^2)^{3/2}}-\frac{2}{\pi a^2}\int_0^\infty \frac{K_0(t)}{I_0(t)}\times tI_0(t\rho/a)\sin(tz/a)dt,$$

$$E_\rho=\frac{\rho}{(z^2+\rho^2)^{3/2}}+\frac{2}{\pi a^2}\int_0^\infty \frac{K_0(t)}{I_0(t)}tI_1(t\rho/a)\cos(tz/a)dt. \qquad (19)$$

It does not seem possible to transform these integrals explicitly into known functions.

Having thus found expressions for the electric field inside the cylinder, we can easily verify them afterwards. For any real value of $\lambda$, the function

$$\cos(\lambda z)\cdot I_0(\lambda\rho)$$

is a regular solution of the potential equation inside the cylinder. Furthermore, the right-hand side of (17), being an absolutely convergent linear combination of those functions, is here a regular potential as well. Therefore, the function $V$ of (18) is also a regular potential, with exception of the origin, where it has just the required singularity. It only remains to be checked that $V$ vanishes on the boundary at $\rho=a$. This, however, is evident from the Fourier inversion of (14):

$$\frac{2}{\pi}\int_0^\infty K_0(\lambda a)\cos\lambda z d\lambda=\frac{1}{(z^2+a^2)^{1/2}}.$$

## 4. A SECOND METHOD TO SOLVE THE POTENTIAL PROBLEM[6]

The *first* method to obtain expressions for the field inside the cylinder, as given above, started from a certain pre-calculated surface charge distribution on the boundary. This method is a little unconventional. Usually one works in the opposite direction: first the potential is obtained as the solution of a boundary value problem, and then the surface charges are calculated. The results above immediately indicate how this *second* method runs. One should try to find a function $f(\lambda)$ such that the potential of the yet unknown charges be expressible as follows

$$V_0(z,\rho)=\int_0^\infty f(\lambda)I_0(\lambda\rho)\cos\lambda z d\lambda,$$

this being a continuous sum of elementary potential functions. Fortunately, the primary potential of the unit point charge can already be expressed in that way:

$$\frac{1}{r}=\frac{2}{\pi}\int_0^\infty K_0(\lambda\rho)\cos\lambda z d\lambda. \qquad (20)$$

and, hence, for the total potential

$$V(z,\rho)=\int_0^\infty\left\{f(\lambda)I_0(\lambda\rho)+\frac{2}{\pi}K_0(\lambda\rho)\right\}\cos\lambda z d\lambda.$$

A sufficient condition that $V(z,\rho)$ be zero on the boundary of the cylinder is that the integrand vanish there; consequently the function $f(\lambda)$ can be determined from

$$f(\lambda)I_0(\lambda a)+\frac{2}{\pi}K_0(\lambda a)=0,$$

and then one obtains (17).

Moreover, in combining Eqs. (17) and (20), we can derive the following expression

$$V(z,\rho)=\frac{2}{\pi}\int_0^\infty\frac{I_0(\lambda a)K_0(\lambda\rho)-K_0(\lambda a)I_0(\lambda\rho)}{I_0(\lambda a)}\times\cos\lambda z d\lambda, \qquad (21)$$

which more directly shows the vanishing at $\rho=a$. Equation (21) enables us to come to the integral expression (5) for the charge density. Theretofore, we use (1) and the Wronskian of $I_0$, $K_0$:

$$I_0(t)K_0'(t)-I_0'(t)K_0(t)=-\frac{1}{t}, \qquad (22)$$

and so we are back at our former starting point.

## 5. APPROXIMATE EXPRESSIONS FOR THE SURFACE CHARGES

Weber,[6] in his theory of a point charge inside a cylinder, follows the second method apart from slight modifications of minor importance. His Eqs. (13), (17), and (18) are equivalent to ours as shown in Eqs. (18) and (19). Equation (5) for

---

[6] E. Weber, "The electrostatic field produced by a point charge in the axis of a cylinder," J. App. Phys. 10, 663 (1939).

the surface charge density, however, was not given by Weber.

As for numerical calculations, Weber writes "the actual evaluation of the field distribution is rather difficult on account of the Fourier integrals which necessitate a point for point numerical integration in infinite limits," and further "the numerical evaluation of the integrals is very laborious." Weber, therefore, does not integrate point for point, but approximates the integrands by more tractable elementary functions, such that the remaining integrals can be computed explicitly.

As a fair approximation, suggested by existing tables of Bessel functions, Weber uses, for instance

$$\frac{K_0(t)}{I_0(t)}tI_1(t)\approx\frac{t}{2}e^{-t},\quad (t>0). \tag{23}$$

From the lower integral in (19), $E_\rho$ can then be evaluated at the surface of the cylinder. Weber's result for the corresponding density of charge, according to $\eta(z)=-E_\rho/4\pi$, may be written

$$-a^2\eta(z)\approx\frac{1}{4\pi}\left\{\frac{1}{(1+\zeta^2)^{3/2}}+\frac{1}{\pi}\frac{1-\zeta^2}{(1+\zeta^2)^2}\right\},$$
$$(\zeta=z/a). \tag{23'}$$

Furthermore, Weber derives expressions for the field in the cut plane as well for points at the surface of the cylinder, by another approximation, different from (23). His approximation (21) reads

$$\frac{K_0(t)}{I_0(t)}\approx(2\pi)^{1/2}te^{-2t},\quad (t>0). \tag{24}$$

We need not, however, confine ourselves to the special values $z=0$ or $\rho=a$, as did Weber. When (24) is applied, we obtain

$$\frac{1}{(2\pi)^{1/2}}\int_0^\infty\frac{K_0(t)}{I_0(t)}I_0(t\rho/a)\cos(tz/a)dt$$
$$=Re\int_0^\infty e^{-t(2+i\zeta)}J_0(it\rho/a)tdt$$
$$\approx Re\frac{2+i\zeta}{[(2+i\zeta)^2-\rho^2/a^2]^{3/2}};$$

TABLE I. Values of surface charge density $|\eta|$ in units $1/a^2$. First column: rigorous values; second: Weber's approximation (23'). Third and fourth columns: approximations (25'), (24'), respectively.

| $z/a$ | $-a^2\eta(z)$ | | | |
|---|---|---|---|---|
| 0 | 0.1055 | 0.1049 | 0.1040 | 0.1285 |
| 0.2 | 0.0980 | 0.0975 | 0.0969 | 0.1144 |
| 0.4 | 0.0795 | 0.0795 | 0.0792 | 0.0827 |
| 0.6 | 0.0581 | 0.0589 | 0.0586 | 0.0510 |
| 0.8 | 0.0396 | 0.0413 | 0.0406 | 0.0280 |
| 1.0 | 0.0259 | 0.0281 | 0.0269 | 0.0147 |
| 1.2 | 0.0165 | 0.0190 | 0.0174 | 0.0076 |
| 1.4 | 0.0104 | 0.0129 | 0.0111 | 0.0041 |
| 1.6 | 0.0065 | 0.0087 | 0.0069 | 0.0025 |
| 1.8 | 0.0040 | 0.0060 | 0.0042 | 0.0018 |
| 2.0 | 0.0025 | 0.0041 | 0.0025 | 0.0016 |

and then, from (18)

$$V(z,\rho)\approx\frac{1}{(z^2+\rho^2)^{1/2}}-\frac{1}{a}\left(\frac{8}{\pi}\right)^{1/2}Re\frac{2+i\zeta}{[(2+i\zeta)^2-\rho^2/a^2]^{3/2}}.$$

Especially on the cut plane, when differentiated to $(-\rho)$, this formula yields

$$E_\rho\approx\frac{1}{\rho^2}\left\{1+\left(\frac{8}{\pi}\right)^{1/2}\frac{6\rho^3/a^3}{[4-\rho^2/a^2]^{5/2}}\right\},\quad (z=0).$$

This result is apparently in disagreement with that of Weber. His relation reads

$$E_{\rho,\text{ Weber}}\approx\frac{1}{\rho^2}\left\{1+\left(\frac{8}{\pi}\right)^{1/2}\frac{\rho^3/a^3}{[4-\rho^2/a^2]^{3/2}}\right\},\quad (z=0).$$

The reason for this discrepancy is that Weber dropped one factor $\gamma$ in the integrand of his integral (22), so that actually he has used the following approximation

$$\frac{K_0(t)}{I_0(t)}\approx(2\pi)^{1/2}e^{-2t} \tag{25}$$

instead of (24).

In the latter case we find for general values of $z$ and $\rho$

$$V(z,\rho)\approx\frac{1}{(z^2+\rho^2)^{1/2}}-\frac{1}{a}\left(\frac{8}{\pi}\right)^{1/2}Re\{(2+i\zeta)^2-\rho^2/a^2\}^{-1/2}$$

leading exactly to Weber's formula for $E_\rho$.

In either of the cases Eqs. (24) and (25), the corresponding surface charges, can be calculated. We obtained, respectively,

$$-a^2\eta(z)\approx\frac{1}{4\pi}\left\{\frac{1}{(1+\zeta^2)^{3/2}}+\left(\frac{8}{\pi}\right)^{1/2}Re\frac{3(2+i\zeta)}{(1+i\zeta)^{5/2}(3+i\zeta)^{5/2}}\right\},\qquad(24')$$

$$-a^2\eta(z)\approx\frac{1}{4\pi}\left\{\frac{1}{(1+\zeta^2)^{3/2}}+\left(\frac{8}{\pi}\right)^{1/2}Re\frac{1}{(1+i\zeta)^{1/2}(3+i\zeta)^{3/2}}\right\}.\qquad(25')$$

Table I shows the three different approximations (23′), (24′), (25′), in presence of the correct values, as computed from (12), (13). The approximation (24′) is obviously very poor, leading to values which are too large at $z\approx 0$ and too small at $z\approx 2a$. Either of the approximations (23′) and (25′) is excellent. That of Weber (23′) is a better approximation at small values of $z$ than is (25′). On the other hand, for values of $z$ exceeding $a$, the latter is much better than the former. Moreover, Weber's formula yields negative values for $-a^2\eta(z)$, when $\zeta>3.6$. This is physically impossible, however. The surface charge density must always have the same sign. The right-hand side of (25′), on the contrary, is positive for large values of $\zeta$. Therefore, as a whole, the approximation (25) may be considered as the most useful one, at least as far as we confine ourselves to the surface charges.

The radial field at the surface of the cylinder, just opposite the point charge, is 1.3262 times as large as it would be under influence of the free-point charge, when the cylinder was removed. The corresponding factors in the three approximations (23), (25), and (24), are 1.3071, 1.3183, and 1.6142, respectively.

## 6. FORMULAS WHICH ADMIT OF NUMERICAL CALCULATIONS INSIDE THE CYLINDER

For exact numerical calculations the integral representation (18) is almost without any value. In Section 2 we derived expressions for the charge density which served its useful purpose. The field can be dealt with in a similar manner. The integral (15), however, is no suitable starting point, because its integrand shows a logarithmic singularity in the lower limit. The difficulties owing to the many-valuedness of the $K$-function may be conveniently avoided in the following way. By partially integrating (15) and using (22), one readily obtains

$$V_0(z,0)=-\frac{2}{\pi z}\int_0^\infty\frac{\sin(tz/a)}{tI_0^2(t)}dt.\qquad(26)$$

We now proceed in the same manner as we did before with (5). In this case the path of integration in the complex plane has to avoid the origin by a small indentation in the upper-half plane. We then obtain for $z>0$

$$\frac{1}{\pi}\int_0^\infty\frac{\sin(tz/a)}{tI_0^2(t)}dt=\tfrac{1}{2}+\sum_{n=1}^{\infty}\underset{(t=ik_n)}{\text{residue}}\frac{e^{itz/a}}{tI_0^2(t)}.$$

We calculate the residue in $t=ik_n$ by putting $t=x+ik_n$, developing the function into ascending powers of $x$, and taking the coefficient of $1/x$ in that expansion. The function is of the type $f(x)/g^2(x)$, where $g(x)$ has a simple zero in $x=0$, and $f(0)\neq 0$. The coefficient of $1/x$ turns out to be

$$\frac{f}{g'^2}\left\{\frac{f'}{f}-\frac{g''}{g'}\right\}\qquad\frac{z}{ak_n}\frac{e^{-k_nz/a}}{J_1^2(k_n)}$$

Consequently, for $z>0$

$$V_0(z,0)=-\frac{1}{z}+\frac{2}{a}\sum_{n=1}^{\infty}\frac{e^{-k_nz/a}}{k_nJ_1^2(k_n)}.$$

To obtain the total potential on the axis of the cylinder we must add the potential of the unit charge in the origin. For positive values of $z$ the latter is $1/z$. Hence

$$V(z,0)=\frac{2}{a}\sum_{n=1}^{\infty}\frac{e^{-k_nz/a}}{k_nJ_1^2(k_n)}.$$

Applying (16), we finally obtain for $z>0$

$$V(z,\rho)=\frac{2}{a}\sum_{n=1}^{\infty}\frac{J_0(k_n\rho/a)}{k_nJ_1^2(k_n)}e^{-k_nz/a}.\qquad(27)$$

The series (27) converges also for $x=0$ with the only exception of the origin, where it diverges to infinity. In actual numerical work (27) appears to be useful for $z/a\geqslant 0.2$; in the remaining range $0\leqslant z/a\leqslant 0.2$ other methods should be applied.

One naturally asks whether or not it is possible to establish developments for the potential, analogous to that of formula (10) for the surface charge density. Actually, the required transformation is readily performed for points in the axis of the cylinder. We only write down the final result:

$$V_0(z,0)=-\frac{2}{\pi z}\sum_{\nu=0}^{\infty}\frac{\sin(\beta\nu+\beta/2)z/a}{(\nu+\frac{1}{2})I_0{}^2(\beta\nu+\beta/2)}$$

$$+\frac{1}{a}\sum_{n=1}^{\infty}\frac{\{1+e^{-2\pi k_n/\beta}\}\cosh(k_nz/a)-(2\pi a/\beta z)\sinh(k_nz/a)}{k_nJ_1{}^2(k_n)\cosh^2(\pi k_n/\beta)}. \quad (28)$$

The first series converges for all real $z$, the second for $|z|\leqslant 2\pi a/\beta$; $\beta$ is an arbitrary positive number. If especially $\beta=2\pi a/z$, the first series vanishes identically, whereas the second one reduces to (27) for $\rho=0$.

For actual numerical evaluation we have chosen $\beta=2$. Then the series (28) becomes

$$-aV_0(z,0)=\frac{a}{z}\{0.79432\sin(z/a)$$
$$+0.01782\sin(3z/a)+0.00034\sin(5z/a)$$
$$+0.00000_6\sin(7z/a)+0.01014\sinh(2.4z/a)\}$$
$$-0.00323\cosh(2.4z/a). \quad (29)$$

Although convergent for $z/a<\pi$, (29) may be useful for $z/a\leqslant 1$.

Unfortunately, there do not seem to exist simple developments for general non-zero values of $\rho$, because (16) yields for the continuation of $(\sin z/a)/(z/a)$ outside the axis

$$\frac{a}{\rho}\int_0^{\rho/a}\cos(tz/\rho)I_0(t)dt,$$

which integral cannot be solved explicitly.

For small values of $z$ we can expand the initial integral (26) into powers of $z/a$. We then obtain

$$V_0(z,0)=--\sum^{\infty}\frac{(-1)^\nu\lambda_{2\nu}}{(2\nu)!}\left(\frac{z}{a}\right)^{2\nu}, \quad (30)$$

where

$$\lambda_{2\nu}=\frac{1}{(2\nu+1)\pi}\int_0^{\infty}\frac{t^{2\nu}}{I_0{}^2(t)}dt, \quad (31)$$

$$\lambda_0=0.4353382,$$
$$\lambda_2=0.2059112,$$
$$\lambda_4=0.6588568.$$

TABLE II. Values of the potential (measured in units $1/a$) inside a grounded cylinder under induction of an axial positive unit-point charge; $a$ = radius of the cylinder, $z$ and $\rho$ the usual cylindrical coordinates. Point charge in the origin.

| $z/a$ \ $\rho/a$ | 0 | 0.2 | 0.4 | 0.6 | 0.8 |
|---|---|---|---|---|---|
| 0 | ∞ | 4.125 | 1.612 | 0.756 | 0.303 |
| 0.1 | 9.131 | 3.600 | 1.539 | 0.736 | 0.297 |
| 0.2 | 4.137 | 2.670 | 1.357 | 0.681 | 0.280 |
| 0.3 | 2.480 | 1.919 | 1.131 | 0.603 | 0.253 |
| 0.4 | 1.661 | 1.395 | 0.916 | 0.518 | 0.224 |
| 0.5 | 1.178 | 1.033 | 0.729 | 0.432 | 0.192 |
| 0.6 | 0.864 | 0.776 | 0.575 | 0.355 | 0.161 |
| 0.7 | 0.648 | 0.591 | 0.410 | 0.287 | 0.133 |
| 0.8 | 0.493 | 0.454 | 0.355 | 0.231 | 0.109 |
| 0.9 | 0.379 | 0.353 | 0.281 | 0.189 | 0.091 |
| 1.0 | 0.292 | 0.272 | 0.219 | 0.147 | 0.071 |
| 1.2 | 0.176 | 0.164 | 0.135 | 0.092 | 0.045 |
| 1.4 | 0.107 | 0.101 | 0.083 | 0.057 | 0.028 |
| 1.6 | 0.066 | 0.062 | 0.051 | 0.036 | 0.017 |
| 1.8 | 0.041 | 0.039 | 0.032 | 0.022 | 0.011 |
| 2.0 | 0.025 | 0.024 | 0.019 | 0.014 | 0.007 |

Except $\lambda_0$, these numerical values were given by Watson.[5]

We can now apply (16) with greater success than before. In performing the transformation (16) upon (30) we obtain

$$V_0(z,\rho)=-\frac{2}{a}\sum_{\nu=0}^{\infty}\frac{(-1)^\nu\lambda_{2\nu}}{(2\nu)!}F_\nu\left(\frac{z}{a},\frac{\rho}{a}\right), \quad (32)$$

where

$$F_\nu(u,v)=\frac{1}{2\pi}\int_0^{2\pi}(u+iv\cos\varphi)^{2\nu}d\varphi$$
$$=(u^2+v^2)^\nu P_{2\nu}\left\{\frac{u}{(u^2+v^2)^{\frac{1}{2}}}\right\}, \quad (33)$$

and $P_{2\nu}$ is the usual symbol for the Legendre polynomial. For instance

$$F_0=1,$$
$$F_1=u^2-\tfrac{1}{2}v^2,$$
$$F_2=u^4-3u^2v^2+\tfrac{3}{8}v^4.$$

It may be remarked that (32) converges for $z^2+\rho^2<4a^2$.

The development (32) becomes much more familiar when spherical coordinates $r$, $\vartheta$, are introduced, connected to the cylindrical ones by means of

$$r^2=z^2+\rho^2;\quad \tan\vartheta=\rho/z.$$

Then (32) is transformed into

$$V_0(z,\rho)=-\frac{2}{a}\sum_{\nu=0}^{\infty}\frac{(-1)^{\nu}\lambda_{2\nu}}{(2\nu)!}\left(\frac{r}{a}\right)^{2\nu}P_{2\nu}(\cos\vartheta). \quad (34)$$

This series is convergent for $r/a<2$; it is nothing else but the development of $V_0(P)$ into spherical potential functions.

Especially in the cut plane $z=0$ we found

$$-aV_0(0,\rho)=0.8707+0.1030(\rho/a)^2+0.0206(\rho/a)^4+0.0057(\rho/a)^6.$$

The first three coefficients are calculated from the values $\lambda_0$, $\lambda_2$, and $\lambda_4$, as specified above; the last coefficient is taken so that the total is 1.0000, on account of $V_0(0,a)$ being $-1/a$, thus neglecting terms of higher order than $(\rho/a)^6$.

Numerical evaluation of the potential inside the cylinder may be readily accomplished now. The different expressions (27), (28), and (34) have reasonable overlapping regions of useful convergence. Numerical data for the total potential are to be found in Table II.

The electric field itself can also be calculated numerically. From (27) we deduce for $z>0$

$$\begin{aligned}E^z&=\frac{2}{a^2}\sum_{n=1}^{\infty}\frac{J_0(k_n\rho/a)}{J_1^2(k_n)}e^{-k_nz/a},\\E_\rho&=\frac{2}{a^2}\sum_{n=1}^{\infty}\frac{J_1(k_n\rho/a)}{J_1^2(k_n)}e^{-k_nz/a}.\end{aligned} \quad (35)$$

It may be noted that in Watson's paper[5] two series $S_1$, $S_2$ occur which are related to the series (35) by

$$E_z=\frac{2}{a^2}S_1;\quad E_\rho=-\frac{2}{a}\frac{\partial S_2}{\partial z}.$$

Therefore, Watson's series can be obtained from that shown in (27). Whereas Watson has given explicitly only the first four terms of the development corresponding to (32), we found the general term here expressible in spherical potential functions; cf. (34). With exclusion of the immediate neighborhood of the cut plane $z=0$, Eqs. (35) are very useful for numerical purposes. Around the cut plane we better use (34) and compute $E_r$, $E_\vartheta$ instead of the cylindrical components; further details may be omitted, however.

## 7. A THIRD METHOD OF SOLUTION

Either of the two distinct methods developed previously yields complete information concerning the field of the axial point charge in the cylinder. The mathematics there involved are quite instructive, and can serve successfully as a guide in problems of wave propagation inside the same cylinder.

The reader is aware that even in the simplest potential problem, as was treated above, a considerable amount of analysis was used. In this respect wave guide theory is still less tractable. The question whether this is really necessary, from an engineer's point of view, will be answered in due course.

In practical wave guide applications, one is mainly interested in the fields far from the exciting antenna, where many of the higher modes have already been damped out. With this restriction in mind, and thus ignoring the field in the nearby zone, we have investigated whether it is possible to calculate the field in the far zone more directly, without the usual extensive amount of complex function theory. This new method was found by inspection and generalization of the expansion (27), obtained for the potential in the preceding section. Each separate term of that series, for non-vanishing $z$, is a solution of the potential equation, fulfilling the required boundary condition at the surface of the cylinder, and showing the suitable behavior at infinity. So we might have expected in advance the possibility of solving our potential problem by means of such a development, in terms of discrete normal solutions of the potential equation:

$$V(z,\rho)=\sum_{n=1}^{\infty}c_nJ_0(k_n\rho/a)e^{-k_n|z|/a}. \quad (36)$$

When the proposed method actually works, it must be possible, somehow, to calculate the unknown coefficients $c_n$ in this development from the behavior of $V(z,\rho)$ in the immediate neighborhood of the prescribed singularity at the origin, although just in the point charge, (36) is known to diverge. In the following we will show how this problem can be solved. We consider a more general problem, however, so that the new method is equally well applicable to solutions of the wave equation. Moreover, we need not restrict ourselves to rotationally symmetrical fields. In many cases the prescribed singularity is of the type

$$\frac{1}{R}=\{z^2+\rho^2-2\rho\rho_0\cos\varphi+\rho_0{}^2\}^{-\frac{1}{2}},$$

where $R$ denotes the distance between the field point $(z, \rho, \varphi)$ and the source at $(0, \rho_0, 0)$. Singularities of this character are commonly used, both in electrostatics and in wave guide theory. We may, however, admit still more general types of sources.

Suppose $G(z, \rho)$ is a *given function* with one or more singularities for $z=0$, $0\leqslant\rho\leqslant a$. Let $G$ be even in $z$, and let it have continuous derivatives of the second order, outside the singularities. In the previous case of an axial point charge, $G=(z^2+\rho^2)^{-\frac{1}{2}}$.

Furthermore, we suppose that $f_n(z)$ is a *given system* $(n=1, 2, \cdots)$ of continuously differentiable functions with the property $f_n'(0)\neq 0$. For instance, $f_n(z)=\exp(-k_nz/a)$.

The problem under consideration can now be stated in the following general form: How to find the coefficients $c_n$ in (37), when the following properties of the function $F(z, \rho)$ are ***known***:

(i) $F$ is an even function of $z$.

(ii) $F-G$ has continuous derivatives of the second order, both outside and inside the singularities of $G$.

(iii) For $z\neq 0$, $F$ can be expanded into a uniformly (in $\rho$) convergent series of the form

$$F(z,\rho)=\sum_{n=1}^{\infty}c_nJ_m(k_{n,m}\rho/a)f_n(|z|), \tag{37}$$

where $k_{n,m}$ denotes the $n$th positive zero of the Bessel function $J_m(k)$, and $m$ is a given integer.

The answer to this question is readily given if appropriate use is made of the well-known Fourier-Bessel-Dini series theory, for details of which the reader may be referred to Watson's standard work on Bessel functions.[7]

On account of the conditions stated under (ii) we infer that $(\partial/\partial z)(F-G)$ can be developed into a Fourier-Bessel series. This series is uniformly convergent in any closed interval, contained in $0<\rho<a$, for any value of $z$, whether outside or on the cut plane $z=0$. Thus we have

$$\frac{\partial}{\partial z}(F-G)=\sum_{n=1}^{\infty}b_n(z)J_m(k_{n,m}\rho/a).$$

The coefficients $b_n(z)$ can be determined in the conventional manner, using the orthogonality of the set of functions $\rho^{\frac{1}{2}}\cdot J_m(k_{n,m}\rho/a)$ upon $0\leqslant\rho\leqslant a$. One finds[7]

$$b_n(z)=\frac{2}{a^2J_{m+1}{}^2(k_{n,m})}\int_0^a\rho\frac{\partial}{\partial z}(F-G)J_m(k_{n,m}\rho/a)d\rho.$$

As either of the two functions $F$, $G$ is even in $z$, with continuous derivatives, we have $b_n(0)=0$. For points *above* the cut plane $(z>0)$ the integral may be broken up into two other integrals, namely

$$b_n(z)=\frac{2}{a^2J_{m+1}{}^2(k_{n,m})}\frac{\partial}{\partial z}\int_0^a\rho F(z,\rho)J_m(k_{n,m}\rho/a)d\rho$$
$$-\frac{2}{a^2J_{m+1}{}^2(k_{n,m})}\int_0^a\rho\frac{\partial G}{\partial z}J_m(k_{n,m}\rho/a)d\rho.$$

From (37), which, for fixed $z\neq 0$ is nothing else but the Fourier-Bessel series of the function $F(z, \rho)$, we obtain

$$\int_0^a\rho F(z,\rho)J_m(k_{n,m}\rho/a)d\rho=c_nf_n(|z|)\frac{a^2}{2}J_{m+1}{}^2(k_{n,m})$$

Therefore, if $z>0$

$$b_n(z)=c_nf_n'(z)-\frac{2}{a^2J_{m+1}{}^2(k_{n,m})}\int_0^a\rho\frac{\partial G}{\partial z}J_m(k_{n,m}\rho/a)d\rho.$$

Now let $z$ tend to zero through positive values; then $b_n(z)\rightarrow 0$, and $f_n'(z)\rightarrow f_n'(0)\neq 0$. Consequently, the above integral must tend to a limit value also, and the coefficient $c_n$ is equal to

$$c_n=\frac{2}{a^2f_n'(0)J_{m+1}{}^2(k_{n,m})}\lim_{z\rightarrow 0^+}\int_0^a\rho\frac{\partial G}{\partial z}\times J_m(k_{n,m}\rho/a)d\rho. \tag{38}$$

The actual evaluation of the limit in (38) usually does not lead to difficulties. Let us take

---

[7] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge, 1944), Chapter XVIII.

as an example the potential problem in its simplest form, whereby the point charge is situated on the axis. Then (38) may be applied for $m=0$, $G=(z^2+\rho^2)^{-\frac{1}{2}}$. In the limit, the main contribution to the integral (38) is attributed to the values of the integrand in the neighborhood of $\rho=0$. Therefore we may substitute unit value for the Bessel function $J_0$. Consequently for the limit value

$$\lim\int_0^a \rho\frac{\partial}{\partial z}(z^2+\rho^2)^{-\frac{1}{2}}d\rho=\lim\frac{\partial}{\partial z}\int_0^a \frac{\rho d\rho}{(z^2+\rho^2)^{\frac{1}{2}}}$$

$$=\lim\frac{\partial}{\partial z}\{(z^2+a^2)^{\frac{1}{2}}-z\}=-1.$$

Furthermore, $f_n'(0)=-k_{n,0}/a=-k_n/a$; thus,

$$c_n=\frac{2}{a^2k_nJ_1^2(k_n)},$$

in agreement with (27).

Until now we were only interested in boundary-value problems of the *first* kind, with vanishing $V$ at the surface of the cylinder. In boundary-value problems of the *second* kind, the normal component of grad $V$ vanishes there; this means $\partial V/\partial\rho=0$ in case of the cylinder. In this second problem the potential can also be developed in terms of Bessel functions. The difference is that the argument $k_{n,m}\rho/a$ is changed into $\kappa_{n,m}\rho/a$, where $\kappa_{n,m}$ means the $n$th positive zero of $J_m'(\kappa)$. Completely analogous to (37), (38), under similar conditions, we now have the following device:

Given

$$\Phi(z,\rho)=\sum_{n=1}^{\infty}\gamma_nJ_m(\kappa_{n,m}\rho/a)f_n(|z|), \qquad (37')$$

"behaves like" $G(z,\rho)$ for $z\rightarrow 0$, then the coefficients $\gamma_n$ can be determined from

$$\gamma_n= a^2f_n'(0)\{1-m^2/\kappa_{n,m}^2\}J_m^2(\kappa_{n,m})$$

$$\times\lim_{z\rightarrow 0^+}\int_0^a \rho\frac{\partial G}{\partial z}J_m(\kappa_{n,m}\rho/a)d\rho. \qquad (38')$$

We would like to emphasize that in practical applications we may forget all the precise mathematical conditions and restrictions. Then there only remains the very simple procedure to obtain the coefficients in the postulated expansions (37) and (37′) from the Eqs. (38) and (38′), respectively. In this connection the actual evaluation of those limits is of secondary importance.

This *third* powerful method is incomplete, in so far that it does not very well provide numerical calculations in the nearby zone, because the convergence of the series (37) and (37′) is usually too slow there. This, however, might always be overcome by suitable extrapolation of the data outside the cut plane. On the other hand, it has many advantages. For instance, unlike the other two methods, it is equally well applicable to either the potential or the wave equation; the only difference occurs in the system of functions $f_n(z)$. Nor does it matter whether or not there is symmetry around the axis of the cylinder.

## 8. FIELD OF AN ECCENTRIC POINT CHARGE

In case of an eccentric point charge, situated at $(0, \rho_0, 0)$, the potential $V(z,\rho,\varphi)$ will be expressible as a Fourier series with respect to $\varphi$, containing only cosine terms. Thus*

$$V(z,\rho,\varphi)=\sum_{m=0}^{\infty}\epsilon_m\cos m\varphi\sum_{n=1}^{\infty}c_{n,m}$$

$$\times J_m(k_{n,m}\rho/a)e^{-k_{n,m}|z|/a}. \qquad (39)$$

The coefficients $c_{n,m}$ have to be determined from the knowledge that $V$ behaves like $(z^2+\rho^2-2\rho\rho_0\cos\varphi+\rho_0^2)^{-\frac{1}{2}}$ in the neighborhood of the source at $(0,\rho_0,0)$. Obviously the method of Section 7 is applicable, if the functions $G$, $F$, are defined as follows:

$$G(z,\rho)\equiv\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\cos m\varphi d\varphi}{(z^2+\rho^2-2\rho\rho_0\cos\varphi+\rho_0^2)^{\frac{1}{2}}}, \qquad (40)$$

$$F(z,\rho)\equiv\frac{1}{2\pi}\int_{-\pi}^{\pi}V(z,\rho,\varphi)\cos m\varphi d\varphi$$

$$=\sum_{n=1}^{\infty}c_{n,m}J_m(k_{n,m}\rho/a)e^{-k_{n,m}|z|/a}. \qquad (40')$$

It remains to show how the limit in (38) can be actually evaluated in case of the special $G$-function (40). This can be accomplished by

---

* $\epsilon_0=1$; $\epsilon_m=2$ if $m>0$.

means of a useful *symbolic* formula, namely

$$\lim_{z\to 0^+}(\partial G/\partial z) = -(1/\rho_0)\delta(\rho-\rho_0), \qquad (41)$$

where $\delta$ denotes a delta-function, "vanishing everywhere, except for $\rho=\rho_0$, such that its integral over $\rho$ from $-\infty$ to $+\infty$ has unit value." As $G$ is even in $z$, one has, outside the singularity, $\partial G/\partial z\to 0$ if $z\to 0$; for $\rho=\rho_0$, this derivative tends to infinity. Therefore, the existence of something like (41) seems reasonable. Our final result is completely rigorous, although (41) has only a symbolic meaning. For further details the reader may be referred to the appendix.

When (41) is applied to (38), we easily obtain

$$c_{n,m} = \frac{2}{a}\frac{J_m(k_{n,m}\rho_0/a)}{k_{n,m}J_{m+1}^2(k_{n,m})},$$

and thus for the potential of the eccentric charge

$$V(z,\rho,\varphi) = \frac{2}{a}\sum_{m=0}^{\infty}\epsilon_m\cos m\varphi$$
$$\times\sum_{n=1}^{\infty}\frac{J_m(k_{n,m}\rho_0/a)J_m(k_{n,m}\rho/a)}{k_{n,m}Y_{m+1}^2(k_{n,m})}e^{-k_{n,m}|z|/a}. \qquad (42)$$

As it should be, this Green function of the first kind for the potential equation inside the cylinder, is symmetrical in $\rho$ and $\rho_0$. For $\rho_0=0$, (42) obviously reduces to (27).

The method of the discrete normal solutions has thus proved its value. We could write the proper solution almost at once. Of course, this is not the only way; it is certainly possible to generalize the first or second method, treated in the preceding sections. For instance, one will find for the analog of (5)

$$\eta(z,\varphi) = -\frac{1}{2\pi^2a^2}\sum_{m=0}^{\infty}\epsilon_m\cos m\varphi$$
$$\times\int_0^{\infty}\frac{I_m(t\rho_0/a)}{I_m(t)}\cos(tz/a)dt,$$

and by contour integration, analogous to (6),

$$\eta(z,\varphi) = -\frac{1}{2\pi a^2}\sum_{m=0}^{\infty}\epsilon_m\cos m\varphi$$
$$\times\sum_{n=1}^{\infty}\frac{J_m(k_{n,m}\rho_0/a)}{J_{m+1}(k_{n,m})}e^{-k_{n,m}|z|/a}.$$

The same expression for the surface charge density is obtained from (42) if (1) is applied to it.

Similarly, other formulas for the axial point charge can be generalized to an eccentric charge. Further details may be omitted. We better demonstrate the usefulness of the third method by a problem of wave propagation.

## 9. THE GREEN FUNCTIONS FOR THE WAVE EQUATION INSIDE THE CYLINDER. RADIATION OF ENERGY BY ACOUSTIC POINT SOURCES

We restrict ourselves to harmonic vibrations of frequency $\omega/2\pi$, and time dependence $\exp(-i\omega t)$. The velocity potential of a free-point source is

$$U_0=e^{ikR}/R;\quad R^2=z^2+\rho^2-2\rho\rho_0\cos\varphi+\rho_0^2. \qquad (43)$$

Outside the point source, the velocity potential $U$ is a solution of the wave equation

$$\Delta U+k^2U=0. \qquad (44)$$

There are two different types of wave functions, according to whether $U$ or $\partial U/\partial\rho$ vanishes at the wall of the cylinder. The Green function of the first kind $U$ is that solution of (44) which vanishes at the surface of the cylinder, which has at infinity the character of plane waves, coming from the direction of the origin of coordinates, and finally such that $U-U_0$ remains finite for $R\to 0$. The Green function of the second kind $U_2$ has a vanishing $\partial/\partial\rho$ at $\rho=a$, the other requirements being the same as those for $U_1$. Instead of the singularity (43) we may take merely $1/R$.

Discrete normal solutions of the wave equation (44) are

$$\cos m\varphi\cdot J_m(k_{n,m}\rho/a)$$
$$\times\exp[-(z/a)(k_{n,m}^2-k^2a^2)^{\frac{1}{2}}], \qquad (45)$$

$$\cos m\varphi\cdot J_m(\kappa_{n,m}\rho/a)$$
$$\times\exp[-(z/a)(\kappa_{n,m}^2-k^2a^2)^{\frac{1}{2}}]. \qquad (45')$$

The functions (45) vanish at $\rho=a$, those of (45′) have a vanishing derivative with respect to $\rho$. On account of symmetry around the plane $\varphi=0$, the potentials do not contain terms with $\sin(m\varphi)$.

Therefore we may expect for $z\neq 0$

$$u_1=\sum_{m=0}^{\infty}\epsilon_m \cos m\varphi \sum_{n=1}^{\infty} c_{n,m}J_m(k_{n,m}\rho/a)$$

$$\times\exp[-(|z|/a)(k_{n,m}^2-k^2a^2)^{\frac{1}{2}}], \quad (46)$$

$$u_2=\sum_{m=0}^{\infty}\epsilon_m \cos m\varphi \sum_{n=1}^{\infty} \gamma_{n,m}J_m(\kappa_{n,m}\rho/a)$$

$$\times\exp[-(|x|/a)(\kappa_{n,m}^2-k^2a^2)^{\frac{1}{2}}]. \quad (46')$$

In these cases the functions $f_n(z)$ are, respectively,

$$\exp[-(z/a)(k_{n,m}^2-k^2a^2)^{\frac{1}{2}}],$$

$$\exp[-(z/a)(\kappa_{n,m}^2-k^2a^2)^{\frac{1}{2}}].$$

It is seen that the only difference between (39) and (46) is in the occurrence of different functions $f_n(z)$. Consequently we can at once write down the result

$$c_{n,m}=\frac{2}{a}\frac{J_m(k_{n,m}\rho_0/a)}{(k_{n,m}^2-k^2a^2)^{\frac{1}{2}}J_{m+1}^2(k_{n,m})}. \quad (47)$$

Obviously the Green function $U_1$ for the wave equation reduces to that of the potential equation (cf. (42)) when $k$ tends to zero. Concerning the square root, we must define for $ka>k_{n,m}$

$$(k_{n,m}^2-k^2a^2)^{\frac{1}{2}}=-i(k^2a^2-k_{n,m}^2)^{\frac{1}{2}}$$

because of the time factor $\exp(-i\omega t)$.

For the second Green function we obtain by application of (38′)

$$\frac{J_m(\kappa_{n,m}\rho_0/a)}{a\,(1-m^2/\kappa_{n,m}^2)(\kappa_{n,m}^2-k^2a^2)^{\frac{1}{2}}J_m^2(\kappa_{n,m})}, \quad (47')$$

It is to be noticed that in case of *resonance* the method fails, because then some coefficient $c$ or $\gamma$ becomes infinitely large. This resonance phenomenon occurs for special values of the applied frequency, namely $ka=k_{n,m}$ or $ka=\kappa_{n,m}$, respectively, for some values of $n$, $m$. They correspond to the cut-off frequencies of the circular wave guide. In practice actual infinities do not occur, because of certain corrections which have been neglected in our treatment, and which only play an important part near the cut-off frequencies.

Once the amplitudes of the different modes of propagation are known, it is not difficult to compute the amount of energy, radiated by the point source into the cylinder. At large distances from the cut plane $z=0$, the potential is determined by the undamped waves only. So we have

$$u_1\approx\sum_{0<k_{n,m}<ka}\sum \epsilon_m c_{n,m}\cos m\varphi J_m(k_{n,m}\rho/a)$$

$$\times\exp[-i(\omega t-(|z|/a)(k^2a^2-k_{n,m}^2)^{\frac{1}{2}})].$$

Because of the fact that the system of functions

$$\cos m\varphi J_m(k_{n,m}\rho/a)$$

is orthogonal over the cross section of the cylinder, one evidently has

$$\frac{1}{2\pi}\iint|u_1|^2dS=\sum|c_{n,m}|^2\int_0^a J_m^2(k_{n,m}\rho/a)\rho d\rho,$$

where the surface integral is extended over the cross section of the cylinder, and the summation extends over $n$, $m$ such that $k_{n,m}<ka$. The integration over $\rho$ can be performed, and after inserting the values of $c_{n,m}$ from (47), one obtains

$$\frac{1}{2\pi}\iint|u_1|^2dS$$

$$=2\sum_{0<k_{n,m}<ka}\frac{J_m^2(k_{n,m}\rho_0/a)}{(k^2a^2-k_{n,m}^2)J_{m+1}^2(k_{n,m})}. \quad (48)$$

Similarly, for the other problem

$$\frac{1}{2\pi}\iint|U_2|^2dS=2\sum_{0<\kappa_{n,m}<ka}$$

$$\times\frac{J_m^2(\kappa_{n,m}\rho_0/a)}{(k^2a^2-\kappa_{n,m}^2)[1-(m^2/\kappa_{n,m}^2)]J_m^2(\kappa_{n,m})}. \quad (48')$$

The physical meaning of the right-hand side of (48′) is that it provides us with the numeric by which the output energy of a *freely* radiating source should be multiplied in order to obtain the energy output, when the same source is placed inside the cylinder, at a distance $\rho_0$ from the axis.

By the method of images we can also treat the case of a point source inside a cylinder with one end closed. Let $d$ denote the distance from the source to the closed end. Then, by introducing an unclosed cylinder with two sources, a distance $2d$ apart, and vibrating out or in phase, we obtain for the potentials

$$u_1 = U_1(z+d,\ \rho,\ \varphi) - U_1(z-d,\ \rho,\ \varphi),$$

$$u_2 = U_2(z+d,\ \rho,\ \varphi) + U_2(z-d,\ \rho,\ \varphi).$$

The far field is a result of two interfering systems of plane waves. It is not difficult to obtain the correspondingly changed energy factors. Before in (48) and (48′) the summation is performed, each separate term must be multiplied by suitable interference factors; these factors are

$$2\sin^2\left[\frac{d}{a}(k^2a^2 - k_{n,m}^2)^{\frac{1}{2}}\right], \tag{49}$$

$$2\cos^2\left[\frac{d}{a}(k^2a^2 - \kappa_{n,m}^2)^{\frac{1}{2}}\right], \tag{49'}$$

respectively. The resulting expressions, which we need not write down in detail, will then provide us with the numerical factors, by which the output of the free source should be multiplied to obtain its output, when placed at distances $d$ and $\rho_0$ from the closed end and the axis, respectively, inside a one-sided cylinder.

Questions of wave propagation in electromagnetic theory can be dealt with in quite the same way, further details of which will not be given at present.

## APPENDIX

The validity of the symbolic formula (41) remains to be proved. That means, in rigorous mathematical terms, we have to prove

$$\lim_{z\to 0^+}\int_0^a \rho f(\rho)\frac{\partial G_m}{\partial z}d\rho = -f(\rho_0), \tag{A}$$

where $G_m$ is defined by

$$G_m = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\cos m\varphi d\varphi}{(z^2+\rho^2-2\rho\rho_0\cos\varphi+\rho_0^2)^{\frac{1}{2}}}. \tag{B}$$

Equation (A) holds true for $0<\rho_0<a$, if $f(\rho)$ is continuous upon the closed interval $(0, a)$, but otherwise arbitrary.

For a complete proof of (A) it is sufficient to know that (A) is valid in the following two special cases:

(1) $f(\rho)=\rho^m$, where $m$ is the same as in $G_m$;

(2) $f(\rho)$ vanishes at $\rho=\rho_0$.

The most general case is obviously a linear combination of these special cases, because one can always write

$$f(\rho) = \frac{f(\rho_0)}{\rho_0^m}\cdot\rho^m + g(\rho),$$

where $g(\rho)$ vanishes at $\rho=\rho_0$.

### Case 1

For $z>0$, from well-known formulas in Bessel-function theory, we have

$$\begin{aligned}(z^2+\rho^2-2\rho\rho_0\cos\varphi+\rho_0^2)^{-\frac{1}{2}} &= \int_0^\infty e^{-zt}J_0[t(\rho^2-2\rho\rho_0\cos\varphi+\rho_0^2)^{\frac{1}{2}}]dt \\ &= \sum_0^\infty \epsilon_\nu \cos\nu\varphi\int_0^\infty e^{-zt}J_\nu(\rho_0t)J_\nu(\rho t)dt.\end{aligned}$$

Upon multiplication by $\cos(m\varphi)$, and integration over a full period of $\varphi$, we obtain

$$G_m = \int_0^\infty e^{-zt}J_m(\rho_0t)J_m(\rho t)dt. \tag{C}$$

Consequently

$$\begin{aligned}\int_0^a \rho\cdot\rho^m(\partial G_m/\partial z)d\rho &= \int_0^a \rho^{m+1}d\rho\int_0^\infty -te^{-zt}J_m(\rho_0t)J_m(\rho t)dt \\ &= -\int_0^\infty t^{-m-1}e^{-zt}J_m(\rho_0t)dt\int_0^{at}\kappa^{m+1}J_m(x)dx \\ &= -a^{m+1}\int_0^\infty e^{-zt}J_m(\rho_0t)J_{m+1}(at)dt.\end{aligned}$$

Let $z$ tend to zero; then the integral above tends to

$$\int_0^\infty J_m(\rho_0t)J_{m+1}(at)dt = (\rho_0^m/a^{m+1}) \quad (0<\rho_0<a).$$

Hence

$$\lim_{z\to 0^+}\int_0^a \rho\cdot\rho^m(\partial G_m/\partial z)d\rho = -\rho_0^m \tag{D}$$

which is (A) for the special function $f(\rho)=\rho^m$.

### Case 2

It is easy to demonstrate that the contribution of the intervals outside the immediate vicinity of $\rho_0$ is zero; that means

$$\lim_{z\to 0^+}\left\{\int_0^{\rho_0-\epsilon}\rho f(\rho)\frac{\partial G_m}{\partial z}d\rho + \int_{\rho_0+\epsilon}^a \rho f(\rho)\frac{\partial G_m}{\partial z}d\rho\right\} = 0, \tag{E}$$

where $\epsilon$ is any positive constant, such that $0\leqslant\rho_0-\epsilon<\rho_0+\epsilon\leqslant a$. For the proof of (E) we introduce $M$, the upper bound of $|f(\rho)|$ upon $(0, a)$. Then

$$\left|\int_0^{\rho_0-\epsilon}\rho f(\rho)\frac{\partial G_m}{\partial z}d\rho\right| \leqslant M\int_0^{\rho_0-\epsilon}\left|\frac{\partial G_m}{\partial z}\right|d\rho.$$

Furthermore,

$$\frac{\partial G_m}{\partial z} = -\frac{z}{2\pi}\int_{-\pi}^{\pi}\frac{\cos m\varphi d\varphi}{(z^2+\rho^2-2\rho\rho_0\cos\varphi+\rho_0^2)^{\frac{3}{2}}};$$

thus

$$\frac{\partial G_m}{\partial z} \leqslant \frac{z}{2\pi}\int_{-\pi}^{\pi}\frac{d\varphi}{(z^2+\rho^2-2\rho\rho_0\cos\varphi+\rho_0^2)^{\frac{3}{2}}} \leqslant \frac{}{[z^2+(\rho-\rho_0)^2]^{\frac{3}{2}}}.$$

Therefore,

$$\left|\frac{\partial G_m}{\partial z}\right| \leqslant \frac{z}{(z^2+\epsilon^2)^{\frac{3}{2}}} \quad \text{if} \quad |\rho-\rho_0|\geqslant\epsilon.$$

Consequently

$$\left|\int_0^{\rho_0-\epsilon}\rho f(\rho)\frac{\partial G_m}{\partial z}d\rho\right| \leqslant \frac{Mz}{(z^2+\epsilon^2)^{\frac{3}{2}}}\int_0^{\rho_0}\rho d\rho.$$

A similar bound is found for the contribution of the interval $(\rho_0+\epsilon, a)$. Upon combining them we obtain

$$\left|\int_0^{\rho_0-\epsilon}\rho f(\rho)\frac{\partial G_m}{\partial z}d\rho+\int_{\rho_0+\epsilon}^{a}\rho f(\rho)\frac{\partial G_m}{\partial z}d\rho\right| \leqslant \frac{Mza^2}{2(z^2+\epsilon^2)^{3/2}}.$$

Let now $z$ tend to zero, $\epsilon$ being kept constant. Then (E) follows at once. (It should be noted that (E) holds for general functions $f(\rho)$, not necessarily vanishing at $\rho_0$.)

For the special function $f(\rho)$, with $f(\rho_0)=0$, the contribution of the remaining interval $\rho_0-\epsilon\leqslant\rho\leqslant\rho_0+\epsilon$ is very small, because $f(\rho)$ is very small there. Let us introduce the upper bound of $|f(\rho)|$ in this interval, thus

$$M(\epsilon)=\max_{|\rho-\rho_0|\leqslant\epsilon}|f(\rho)|,$$

then obviously $M(\epsilon)\rightarrow 0$ if $\epsilon\rightarrow 0$. Furthermore,

$$\left|\frac{\partial G_m}{\partial z}\right| \leqslant \frac{z}{2\pi}\int_{-\pi}^{+\pi}\frac{d\varphi}{(z^2+\rho^2-2\rho\rho_0\cos\varphi+\rho_0^2)^{3/2}}=-\frac{\partial G_0}{\partial z}.$$

Therefore,

$$\left|\int_{\rho_0-\epsilon}^{\rho_0+\epsilon}\rho f(\rho)\frac{\partial G_m}{\partial z}d\rho\right| \leqslant -M(\epsilon)\int_{\rho_0-\epsilon}^{\rho_0+\epsilon}\rho\frac{\partial G_0}{\partial z}d\rho$$

$$\leqslant -M(\epsilon)\int_0^a\rho\frac{\partial G_0}{\partial z}d\rho.$$

Let now $z$ tend to zero; then the right-hand integral tends to $-1$, as follows from (D) when the latter formula is applied with $m=0$.

Thus we have

$$\left|\lim_{z\rightarrow 0+}\int_{\rho_0-\epsilon}^{\rho_0+\epsilon}\rho f(\rho)\frac{\partial G_m}{\partial z}d\rho\right| \leqslant M(\epsilon);$$

that means, on account of (E)

$$-M(\epsilon)\leqslant\lim_{z\rightarrow 0+}\int_0^a\rho f(\rho)\frac{\partial G_m}{\partial z}d\rho\leqslant M(\epsilon).$$

This is true for any $\epsilon>0$; as $M(\epsilon)\rightarrow 0$ for $\epsilon\rightarrow 0$, consequently

$$\lim_{z\rightarrow 0+}\int_0^a\rho f(\rho)\frac{\partial G_m}{\partial z}d\rho=0=-f(\rho_0)$$

and this is the required expression in case 2. So we have rigorously proved the validity of the symbolic formula (41). It may be noticed here that according to Watson[8] we can deduce from (C) the following interesting formula

$$G_m=\frac{1}{\pi(\rho\rho_0)^{1/2}}Q_{m-1/2}\left(\frac{z^2+\rho^2+\rho_0^2}{2\rho\rho_0}\right), \tag{F}$$

where $Q$ denotes the Legendre function of the second kind. The $Q$-function has a singularity at the point where the argument is equal to 1; this corresponds with $z=0$, $\rho=\rho_0$.

Eindhoven, March, 1946.

[8] G. N. Watson, reference 7, page 389, Eq. (2).

# The Elastic Constants of Materials Loaded with Non-Rigid Fillers

JANE M. DEWEY
*General Laboratories, United States Rubber Company, Passaic, New Jersey*

(Received November 14, 1946)

General expressions are given for the modulus of rigidity, $\bar{\lambda}_2$, and compressibility, $\bar{k}$, of a medium of Lame's constants, $\lambda_1$ and $\lambda_2$, loaded with a volume fraction, $\phi$, of a filler of constants $\lambda_1'$ and $\lambda_2'$. To terms in the first power of $\phi$

$$\bar{\lambda}_2=\lambda_2\left(1+\frac{15(\lambda_2'-\lambda_2)(\lambda_1+2\lambda_2)}{2\lambda_2'(3\lambda_1+8\lambda_2)+\lambda_2(9\lambda_1+14\lambda_2)}\phi\right)$$

$$\bar{k}=k+\frac{3+4\lambda_2k'}{3+4\lambda_2k}(k'-k)\phi.$$

If the "filler" is a gas at a pressure, $p$, in a nearly incompressible medium, these give

Young's modulus $\quad \bar{E}=E\left(1-\frac{E}{9p+4E}\phi\right).$

Modulus of rigidity $\quad \bar{\lambda}_2=\lambda_2(1-\frac{5}{3}\phi).$

Compressibility $\quad \bar{k}=k+\frac{3}{3p+4\lambda_2}\phi.$

Poisson's ratio $\quad \bar{\sigma}=\sigma\left(1-\frac{3E}{9p+4E}\phi\right),$

where $\phi$ is the volume loading, $p$ is the pressure within the spherical cavities in the deformed state. Barred symbols refer to the properties of the loaded material; unbarred, to the medium alone. Expressions for the displacements and stresses within the medium and the particles, neglecting interactions between particles, are also given.

---

IN a letter to the Editor of this Journal,[1] the writer gave a solution of the linear equations describing small elastic deformations of a medium containing a single spherical inclusion of any elastic constants. This problem had previously been solved for simple tension by J. N. Goodier.[2] The solution can be extended to the determination of the elastic constants of a medium loaded with small spheres by the method applied by A. Einstein[3] to the calculation of the viscosity of a suspension and by H. M. Smallwood[4] to the problem of rigid fillers. This method gives the terms in the first power of the volume loading only, and is valid, with neglect of surface effects, for very dilute suspensions.[*, 5]

Consider a single sphere of radius $R$ at the origin of coordinates in an isotropic medium in which the displacement, $\mathbf{s}$, at a large distance from the sphere is given by

$$\mathbf{s}\underset{r\to\infty}{\longrightarrow}Axi+Byj+Czk. \tag{1}$$

Lame's constants of the sphere will be denoted by $\lambda_1'$ and $\lambda_2'$, of the medium by $\lambda_1$ and $\lambda_2$. $\lambda_2$ is the modulus of rigidity, $\lambda_1$ may be defined by $\lambda_1+\frac{2}{3}\lambda_2=k^{-1}$, where $k$ is the compressibility.

The dilatation, strain, and radial components of the stress tensor, $\mathbf{T}$, in the medium are

$$\nabla\cdot\mathbf{s}=A+B+C\\ -6\lambda_2(3\lambda_1+5\lambda_2)^{-1}c_{-2}r^{-3}M_1(\theta,\phi), \tag{2}$$

$$s_r=\tfrac{1}{3}(r+b_{-2}r^{-2})(A+B+C)\\ +(r+c_{-2}r^{-2}+c_{-4}r^{-4})M_1(\theta,\phi), \tag{3}$$

$$s_\theta=[r+2\lambda_2(3\lambda_1+5\lambda_2)^{-1}c_{-2}r^{-2}\\ -\tfrac{2}{3}c_{-4}r^{-4}]M_2(\theta,\phi), \tag{4}$$

$$s_\phi=s_\theta M_2^{-1}(\theta,\phi)M_3(\theta,\phi), \tag{5}$$

---

[1] Jane M. Dewey, J. App. Phys. **16**, 55 (1945); Erratum, **18**, 132 (1947).

[2] J. N. Goodier, J. App. Mech. **1**, 39 (1933).

[3] A. Einstein, Ann. d. Physik **19**, 289 (1906) and **34**, 591 (1911).

[4] H. M. Smallwood, J. App. Phys. **15**, 758 (1944).

[*] The numerical value of the coefficient of the terms in the second power of the volume loading is given by Guth and Gold (reference 5), but no details of their calculation are available.

[5] E. Guth and O. Gold, Phys. Rev. **53**, 322 (1938) (abstract). See also E. Guth, J. App. Phys. **16**, 20 (1945).

$$T_{rr}=(\lambda_1+\tfrac{2}{3}\lambda_2-\tfrac{4}{3}\lambda_2 b_{-2}r^{-3})(A+B+C)$$
$$+2\lambda_2[1-(9\lambda_1+10\lambda_2)(3\lambda_1+5\lambda_2)^{-1}$$
$$\times c_{-2}r^{-3}-4c_{-4}r^{-5}]M_1(\theta,\phi) \quad (6)$$

$$T_{r\theta}=2\lambda_2[1+(3\lambda_1+2\lambda_2)$$
$$\times(3\lambda_1+5\lambda_2)^{-1}c_{-2}r^{-3}$$
$$+\tfrac{8}{3}c_{-4}r^{-5}]M_2(\theta,\phi), \quad (7)$$

$$T_{r\phi}=T_{r\theta}M_2^{-1}(\theta,\phi)M_3(\theta,\phi), \quad (8)$$

$$M_1(\theta,\phi)=A\sin^2\theta\cos^2\phi+B\sin^2\theta\sin^2\phi$$
$$+C\cos^2\theta-\tfrac{1}{3}(A+B+C), \quad (9)$$

$$M_2(\theta,\phi)=\tfrac{1}{2}\frac{\partial M_1}{\partial\theta}=\sin\theta\cos\theta$$
$$\times(A\cos^2\phi+B\sin^2\phi-C), \quad (10)$$

$$M_3(\theta,\phi)=\frac{1}{2\sin\theta}\frac{\partial M_1}{\partial\phi}$$
$$=-(A-B)\sin\theta\sin\phi\cos\phi, \quad (11)$$

$$b_{-2}=(3\lambda_1+2\lambda_2-3\lambda_1'-2\lambda_2')$$
$$\times(4\lambda_2+3\lambda_1'+2\lambda_2')^{-1}R^3, \quad (12)$$

$$c_{-2}=-5(\lambda_2'-\lambda_2)(3\lambda_1+5\lambda_2)$$
$$\times[2\lambda_2'(3\lambda_1+8\lambda_2)$$
$$+\lambda_2(9\lambda_1+14\lambda_2)]^{-1}R^3, \quad (13)$$

$$c_{-4}=9(\lambda_2'-\lambda_2)(\lambda_1+\lambda_2)$$
$$\times[2\lambda_2'(3\lambda_1+8\lambda_2)$$
$$+\lambda_2(9\lambda_1+14\lambda_2)]^{-1}R^5. \quad (14)$$

The energy of deformation is given by

$$2W=\int_S \mathbf{T}_n\cdot\mathbf{s}da, \quad (15)$$

where $\mathbf{T}_n$ is the normal component of the stress tensor and the integration is over the outer surface of the medium. To calculate the energy of deformation of an infinite medium containing a single imbedded sphere, it is convenient to take a spherical surface with the particle at the center. Then,

$$W=\tfrac{1}{2}\int_S (T_{rr}s_r+T_{r\theta}s_\theta+T_{r\phi}s_\phi)da,$$
$$=\tfrac{1}{2}V[\lambda_1(A+B+C)^2$$
$$+2\lambda_2(A^2+B^2+C^2)]+\frac{2\pi}{3}(\lambda_1-\tfrac{2}{3}\lambda_2)$$
$$\times\{[b_{-2}-\tfrac{2}{5}\lambda_2(3\lambda_1+5\lambda_2)^{-1}c_{-2}]$$
$$\times(A+B+C)^2-2\lambda_2c_{-2}(3\lambda_1+5\lambda_2)^{-1}$$
$$\times(A^2+B^2+C^2)\}, \quad (16)$$

for an infinite volume, where $V$ is the total volume of the loaded medium.

Assuming that to the first power of the volume loading the added energy due to $N$ spheres is $N$ times the added energy due to a single sphere

$$W/V=\tfrac{1}{2}\{\lambda_1+(\lambda_1-\tfrac{2}{3}\lambda_2)$$
$$\times[b_{-2}-\tfrac{2}{5}\lambda_2(3\lambda_1+5\lambda_2)^{-1}c_{-2}]R^{-3}\phi\}$$
$$\times(A+B+C)^2+\lambda_2[1-(\lambda_1-\tfrac{2}{3}\lambda_2)$$
$$\times(3\lambda_1+5\lambda_2)^{-1}c_{-2}R^{-3}\phi]$$
$$\times(A^2+B^2+C^2), \quad (17)$$

where $\phi=\tfrac{4}{3}\pi NR^3/V$, the partial volume of the filler particles.

The strain constants, $A$, $B$, and $C$ of this expression are those for the medium, as appears from the derivation. The observed mean strain constants, $A_0$, $B_0$, and $C_0$, can be calculated to the first approximation by the method outlined by Einstein. The displacement in the $x$ direction can be written

$$s_x=Ax+\textstyle\sum_i \Delta s_{xi}=A_0x, \quad (18)$$

where $\Delta s_{xi}$ is the added displacement due to the $i$th particle.

If $X_i$ is the $x$ coordinate of the center of the $i$th particle and $x_i$ is the $x$ coordinate measured from $X_i$

$$x=x_i+X_i. \quad (19)$$

Differentiating (18)

$$A_0=A+\frac{\partial}{\partial x}\textstyle\sum_i \Delta s_{xi}=A+\sum_i\frac{\partial}{\partial x_i}\Delta s_{xi}. \quad (20)$$

Assume that the added displacement is given by Eqs. (3) to (5) less the displacement in the medium at a distance from a particle

$$\Delta s_{xi}=s_{xi}-Ax=s_{xi}-Ax_i-AX_i. \qquad (21)$$

Two assumptions are made here: the first, that at any point the added displacement due to each particle is the same as if the other particles were not present, and second, that the displacement at a distance from a particle is described by the boundary condition of Eq. (1). The first assumption should be correct to the first approximation in the volume loading. The second is approximately correct if each particle is surrounded by a region free of particles of dimensions much greater than the particle radius. This is true of most of the particles if they are randomly distributed and the volume loading is much less than one.

For randomly distributed particles $\sum\Delta X_i=0$ and

$$\sum \Delta s_{xi}=\sum s_{xi}-Ax_i. \qquad (22)$$

We replace the summation by $N$ times the mean value of $\Delta s_{xi}$

$$\sum_i \frac{\partial}{\partial x_i}\Delta s_{xi}=N/V\int_V \frac{\partial \Delta s_{xi}}{\partial x_i}dV=N/V\int_S \Delta s_{xi}\frac{x_i}{r_i}da, \qquad (23)$$

where the first integral is over the volume and the second over the surface. To terms in $R^3r^{-2}$ the integral is independent of $r_i$ and since the limits of angle are independent of $i$,

$$A_0=A+N/V\int_S \Delta s_x \sin\theta\cos\phi da, \qquad (24)$$

where $\Delta s_x$ may be taken for a particle at the origin. $B_0$ and $C_0$ are obtained by interchanging variables.

Carrying out the integration

$$\begin{aligned}A_0=&(1+\tfrac{2}{5}(3\lambda_1+8\lambda_2)(3\lambda_1+5\lambda_2)^{-1}R^{-3}\phi)A\\&+\tfrac{1}{3}[b_{-2}R^{-3}-\tfrac{2}{5}(3\lambda_1+8\lambda_2)\\&\times(3\lambda_1+5\lambda_2)^{-1}c_{-2}R^{-3}](A+B+C)\phi, \qquad (25)\end{aligned}$$

whence, by combining with corresponding expressions for $B_0$ and $C_0$,

$$A_0+B_0+C_0=(1+b_{-2}\phi)(A+B+C), \qquad (26)$$

$$\begin{aligned}A_0^2+B_0^2+C_0^2=&[1+\tfrac{4}{5}(3\lambda_1+8\lambda_2)\\&\times(3\lambda_1+5\lambda_2)^{-1}c_{-2}R^{-3}\phi]\\&\times(A^2+B^2+C^2)\\&+\tfrac{2}{3}(b_{-2}-\tfrac{2}{5}c_{-2})R^{-3}\\&\times(A+B+C)^2\phi. \qquad (27)\end{aligned}$$

The constants $\bar{\lambda}_1$ and $\bar{\lambda}_2$ of the suspension are obtained by equating the expression for the work per unit volume

$$\begin{aligned}W/V=&\tfrac{1}{2}\bar{\lambda}_1(A_0+B_0+C_0)^2\\&+\bar{\lambda}_2(A_0^2+B_0^2+C_0^2) \qquad (28)\end{aligned}$$

to the expression (17).

Substituting (26) and (27) in (17) and equating coefficients of $(A_0+B_0+C_0)^2$

$$\begin{aligned}\bar{\lambda}_1=&\lambda_1-(\lambda_1+2\lambda_2)\\&\times(b_{-2}-2\lambda_2(3\lambda_1+5\lambda_2)^{-1}c_{-2})R^{-3}\phi. \qquad (29)\end{aligned}$$

Equating coefficients of $(A_0^2+B_0^2+C_0^2)$,

$$\begin{aligned}\bar{\lambda}_2=&\lambda_2[1-3(\lambda_1+2\lambda_2)(3\lambda_1+5\lambda_2)^{-1}c_{-2}R^{-3}\phi]\\=&\lambda_2\{1+15(\lambda_2'-\lambda_2)(\lambda_1+2\lambda_2)\\&\times[2\lambda_2'(3\lambda_1+8\lambda_2)\\&+\lambda_2(9\lambda_1+14\lambda_2)]^{-1}\phi\}. \qquad (30)\end{aligned}$$

Whence

$$\frac{1}{\bar{k}}=\bar{\lambda}_1+\tfrac{2}{3}\bar{\lambda}_2=\frac{1}{k}[1-(\lambda_1+2\lambda_2)b_{-2}R^{-3}\phi], \qquad (31)$$

or

$$k=k+(3+4\lambda_2k)(3+4\lambda_2k')^{-1}(k'-k)\phi. \qquad (32)$$

Two special cases are of importance:

(a) The case of a rigid filler,

$$\lambda_2'\gg\lambda_2, \quad \text{and} \quad k'\ll k.$$

Then,

$$k=k[1-(1+\tfrac{4}{3}\lambda_2k)\phi]. \qquad (32a)$$

$$\bar{\lambda}_2=\lambda_2[1+\tfrac{5}{2}(\lambda_1+2\lambda_2)(\lambda_1+\tfrac{8}{3}\lambda_2)^{-1}\phi]. \qquad (30a)$$

(b) The case of included air or other gas

$$\lambda_2'=0, \quad \lambda_1'\simeq p,$$

where $p$ is the pressure of the included gas

(after the deformation)*

$$k = k + (3+4\lambda_2 k)(3p+4\lambda_2)^{-1}(1-kp)\phi. \quad \text{(32b)}$$

The application of this formula to data on elastomers containing small amounts of gas is difficult as $p$ is not known and $\lambda_2$ and $p$ are of the same order of magnitude. If aged samples of reasonably permeable elastomers are studied, $p$ can probably be taken as 1 atmosphere. In GR-I or other nearly impermeable materials, $p$, for very long times after molding, is probably approximately equal to the gas pressure at the time of molding. It may, therefore, differ considerably from 1 atmosphere. For any reasonably low gas pressure $kp \ll 1$ and

$$\bar{k} \simeq k + (3+4\lambda_2 k)(3p+4\lambda_2)^{-1}\phi, \quad \text{(32b')}$$

$$\bar{\lambda}_2 \simeq \lambda_2[1 - \tfrac{5}{3}(\lambda_1+2\lambda_2)(\lambda_1+\tfrac{14}{9}\lambda_2)^{-1}\phi]. \quad \text{(30b)}$$

The approximate formulae for Young's modulus, $E$, and Poisson's ratio, $\sigma$, valid for a medium of $\lambda_1 \gg \lambda_2$, containing gas at a low pressure are

$$\bar{E} \simeq E[1 - E(9p+4E)^{-1}\phi], \quad \text{(33)}$$

$$\bar{\sigma} \simeq \sigma[1 - 3E(9p+4E)^{-1}\phi]. \quad \text{(34)}$$

The percentage changes in the modulus of rigidity and Young's modulus are of the same order of magnitude as the volume percent of gas in the material. The change in compressibility of a gum rubber stock caused by 0.01 percent of gas at atmospheric pressure is of the order of the compressibility, however. At pressures above an atmosphere, the compressibility varies rapidly with the pressure of the included gas.

Within an embedded sphere and at its surface,

$$\nabla \cdot \mathbf{s} = 3a_1, \quad \text{(35)}$$

$$s_r = (c_1 M_1(\theta, \phi) + a_1)r, \quad \text{(36)}$$

$$s_\theta = c_1 r M_2(\theta, \phi), \quad \text{(37)}$$

$$s_\phi = c_1 r M_3(\theta, \phi), \quad \text{(38)}$$

$$T_{rr} = 2\lambda_2' c_1 M_1(\theta, \phi) + (3\lambda_1' + 2\lambda_2')a_1, \quad \text{(39)}$$

$$T_{r\theta} = 2\lambda_2' c_1 M_2(\theta, \phi), \quad \text{(40)}$$

$$T_{r\phi} = 2\lambda_2' c_1 M_3(\theta, \phi), \quad \text{(41)}$$

$$a_1 = (\lambda_1 + 2\lambda_2)(4\lambda_2 + 3\lambda_1' + 2\lambda_2')^{-1} \times (A+B+C), \quad \text{(42)}$$

$$c_1 = 15\lambda_2(\lambda_1+2\lambda_2)[2\lambda_2'(3\lambda_1+8\lambda_2) + \lambda_2(9\lambda_1+14\lambda_2)]^{-1}. \quad \text{(43)}$$

As would be expected, the radial components of stress on the surface of a particle have their maximum value when the filler is rigid, the case treated by Smallwood.[4]

Comparison of Eq. (30) for the modulus of rigidity with the calculation made by Taylor[6] of the viscosity of a suspension of fluid spheres of viscosity in a fluid of viscosity

$$\bar{\mu} = \mu\left[1 + \frac{5\mu' + 2\mu}{2(\mu' + \mu)}\phi\right],$$

shows that the two relations are of the same form only for $\lambda_1$, $\lambda_1'$, $\mu'$, and $\lambda_2'$ infinite. The difference between the two cases lies in the importance of surface tension in the suspension of fluid droplets. Suspension of droplets of any fluid in another increases the viscosity, while blowing, for instance, reduces the rigidity of a solid. Added work on the fluid suspension is done against surface tension, deforming the droplets. Mathematically, the difference appears in the omission of the condition that the normal component of stress on the droplet be single valued. This difference in stress on the two sides of the surface is balanced by surface tension.

---

* The equations and boundary conditions used in determining the displacements and stresses of Eqs. (2)-(8) are the conditions of equilibrium for a media obeying Hooke's law and a linear relation between volume and pressure. When differential values of the constants are used, these conditions are valid for media of any type, since a linear relation may always be assumed for an infinitesimal deformation. For an ideal fluid $\lambda_2 = 0$ and $dp = -\lambda_1(dV/V)$, which reduces approximately to $\lambda_1 = p$ for a gas.

[6] G. I. Taylor, Proc. Roy. Soc. **A138**, 41 (1932).

# Heat Transfer between a Fluid and a Porous Solid Generating Heat[1]

Stuart R. Brinkley, Jr.[2]
*Central Experiment Station, U. S. Bureau of Mines, Pittsburgh, Pennsylvania*

(Received September 16, 1946)

The theory, due to Anzelius, of the transfer of heat between a fluid flowing with constant velocity through a porous solid is extended to include the case where the solid generates heat. Expressions are obtained for the temperatures of solid and fluid as functions of position and time, it being assumed that the heat source function is a linear function of the temperature of the solid with coefficients independent of position and time. The application of the theory to a description of the temperature in catalytic reactions is indicated.

## INTRODUCTION

A THEORY of the transfer of heat between a porous body and a warm fluid has been developed by Anzelius[3] and Schumann.[4] A fluid carrying heat is supposed to flow with constant velocity through a porous substance such as a solid body in a finely divided state which is initially at a different temperature than that of the fluid. The temperatures of solid and fluid are determined for different positions at later times, it being assumed that the volume specific heats of solid and fluid are constant, and that there is no conduction of heat in either the solid or the fluid. The problem is, of course, idealized by these simplifying assumptions. However, in his extensive series of measurements of heat transfer from a gas stream to a bed of broken solids, Furnas[5] obtained excellent experimental confirmation of the simple theory.

In the present communication, this theory is extended to the case where the solid is generating heat. In addition to the assumptions above, it is further assumed that the heat source function is a linear function of the temperature of the porous solid, the parameters of the linear function being independent of position or time.

The theory should be applicable to a calculation of solid and gas temperatures in catalytic reactions if the reaction vessel is so designed that all but a negligible amount of the heat of reaction is transported by the sensible heat of the gas and if either (a) the reaction is of the zeroth other without volume change, or (b) the flux of the gas through the reaction vessel is so large that reaction is limited to a few percent.

## THE BASIC EQUATIONS

Consider a semi-infinite ($x \geqslant 0$) porous solid through which a fluid is flowing in the positive $x$-direction with volume current $g$. Let the temperatures of solid and fluid at a distance $x$ from the origin and at time $t$ be $U(x, t)$ and $V(x, t)$, respectively. It is assumed that the solid in a particular element of volume loses heat by transfer to the fluid and gains heat by generation at a rate $F$ per unit volume of porous solid, where in general,

$$F = F(x, t; U), \tag{1}$$

and that the fluid in the element of volume gains heat by transfer from the solid and by the flux of heat into the volume which accompanies the flow of the fluid. By considering the rate of increase of heat in the solid and fluid contained between the transverse planes $x$ and $x+dx$, and assuming that the rate of heat transfer is governed by Newton's law of cooling, it is easy to show that the transfer of heat from solid to fluid is governed by the relations,

$$\begin{gathered} c_s\rho_s\frac{\partial U}{\partial t}+h(U-V)-F(x, t; U)=0, \\ c_f\rho_f\frac{\partial V}{\partial t}+h(V-U)+c_f\rho_f g\frac{\partial V}{\partial x}=0, \end{gathered} \tag{2}$$

where $c_s$ and $c_f$ are the heat capacities per unit

---

[1] Published by permission of the Director, Bureau of Mines, U. S. Department of the Interior.

[2] Physical chemist, Research and Development Division, Bureau of Mines, Pittsburgh, Pennsylvania.

[3] A. Anzelius, Zeits. f. ang. Math. u. Mech. **6**, 291 (1926).

[4] T. E. W. Schumann, J. Frank. Inst. **208**, 405 (1929).

[5] C. C. Furnas, U. S. Bureau of Mines Bulletin 361 (1932).

mass of solid and fluid, respectively, $\rho_s$ and $\rho_f$ are masses of solid and gas contained in unit volume of porous solid, respectively, $h$ is the coefficient of heat transfer between solid and gas per unit volume of porous solid, and where it has been assumed that the quantities $c_s\rho_s$, $c_f\rho_f$, and $g$ are constants.

In the present communication, the solution of Eqs. (2) is obtained for the case where the heat source function is a linear function of the temperature of the solid,

$$F(x, t;\, U)=h(\alpha+\beta U), \tag{3}$$

with $\alpha$ and $\beta$ constant. If the heat is generated by chemical reaction, one will usually have

$$F\sim e^{-\epsilon/RT}=e^{-\epsilon/RT_0}\left[1+\frac{\epsilon}{RT_0^2}(T-T_0)+\cdots\right],$$

from which the accuracy of the linear approximation, Eq. (3), may be estimated. Here, $\epsilon$ is the activation energy of the reaction, $R$ is the gas constant, and $T$ is the absolute temperature.

For convenience, the temperature of the inlet fluid is taken to be zero, and the temperature of the porous solid at the initial instant of time on the same temperature scale is denoted by $U_0$. The boundary conditions for Eqs. (2) are therefore

$$\begin{aligned} &U=U_0, \quad x\geqslant gt,\\ &V=0, \quad\;\; x=0. \end{aligned} \tag{4}$$

The first condition is a statement of the fact that the porous solid is at its initial temperature at points which have not yet been reached by the fluid.

With the notation $a=h/c_f\rho_f g$, $b=h/c_s\rho_s g$, and the change of independent variable

$$\xi=ax, \quad \tau=b(gt-x), \tag{5}$$

Eqs. (2) and (3) become

$$\begin{aligned} \partial U/\partial\tau&=V+(\beta-1)U+\alpha,\\ \partial V/\partial\xi&=U-V, \end{aligned} \tag{6}$$

with the boundary conditions, $U=U(\xi,\tau)$, $V=V(\xi,\tau)$,

$$U(\xi, 0)=U_0, \quad V(0, \tau)=0. \tag{7}$$

Equations (6) and (7) lead immediately to a pair of ordinary differential equations for $U(0,\tau)$ and $V(\xi, 0)$.

$$\begin{aligned} &\frac{d}{d\tau}U(0,\tau)+(1-\beta)U(0,\tau)=\alpha, \quad U(0,0)=U_0,\\ &\frac{d}{d\xi}V(\xi,0)+V(\xi,0)=U_0, \quad V(0,0)=0. \end{aligned} \tag{8}$$

The subsidiary boundary conditions are obtained by solution of Eq. (8).

$$\begin{aligned} &U(0,\tau)=U_0e^{(\beta-1)\tau}-\frac{\alpha}{1-\beta}\left[e^{(\beta-1)\tau}-1\right],\\ &V(\xi,0)=U_0(1-e^{-\xi}). \end{aligned} \tag{9}$$

Equation (6) may be simplified in a manner analogous to that employed by Schumann[4] by introducing new independent variables, $\Gamma$ and $\Delta$, where

$$\begin{aligned} \Gamma&=(U+V)e^{\xi+(1-\beta)\tau},\\ \Delta&=(U-V)e^{\xi+(1-\beta)\tau}. \end{aligned} \tag{10}$$

One obtains

$$\begin{aligned} &\partial\Gamma/\partial\tau+\partial\Delta/\partial\tau=\Gamma-\Delta+2\alpha e^{\xi+(1-\beta)\tau},\\ &\partial\Gamma/\partial\xi-\partial\Delta/\partial\xi=\Gamma+\Delta, \end{aligned} \tag{11}$$

and by further differentiation

$$\partial^2\Delta/\partial\xi\partial\tau=\Delta. \tag{12}$$

The boundary conditions for $\Gamma$ and $\Delta$ are evidently

$$\begin{aligned} &\Gamma(\xi,0)=U_0(2e^{\xi}-1), \quad \Delta(\xi,0)=U_0,\\ &\Gamma(0,\tau)=\Delta(\xi,0)=U_0+\frac{\alpha}{1-\beta}\left[e^{(1-\beta)\tau}-1\right]. \end{aligned} \tag{13}$$

The characteristics of Eq. (12) are the straight lines parallel to the axes through the point of interest $(\xi, \tau)$, and the boundary values of $\Delta$ are specified along a pair of characteristics (the axes). The solution of Eqs. (12) and (13) is thus a characteristic initial value problem and is straightforward by the method of Riemann.[6] The Reimann function is $I_0[2((\xi-\xi')(\tau-\tau'))^{\frac{1}{2}}]$, in which $\xi'$ and $\tau'$ are the variables of integration around the boundary of the region determined by the characteristics and the axes, and $I_0$ is the zeroth-order Bessel function of the first kind for

---

[6] R. Courant and D. Hilbert, *Methoden der Mathematischen Physik* (Julius Springer, Berlin 1924), Vol. II, pp. 311–317.

imaginary argument. The solution for $\Delta$ is

$$\Delta = U_0 I_0[2(\xi\tau)^{\frac{1}{2}}] + \alpha \int_0^{\tau} e^{(1-\beta)\tau'} I_0[2(\xi(\tau-\tau'))^{\frac{1}{2}}] d\tau'. \qquad (14)$$

This result may be rewritten in the convenient form,

$$\Delta = U_0 I_0[2(\xi\tau)^{\frac{1}{2}}] + \frac{\alpha}{1-\beta} \varphi[\tau(1-\beta), \xi/(1-\beta)], \qquad (15)$$

in which the function $\varphi(x, y)$, defined by

$$\varphi(x, y) = e^{x} \int_0^{x} e^{-x'} I_0[2(x'y)^{\frac{1}{2}}] dx', \qquad (16)$$

has been introduced. This function, originated by Laplace,[7] has been discussed by Thomas,[8] who has obtained the following useful properties:

$$\left.\begin{aligned} &\varphi(x, y) + \varphi(y, x) = e^{x+y} - I_0[2(xy)^{\frac{1}{2}}], \\ &(\partial/\partial x)\varphi(x, y) = \varphi(x, y) + I_0[2(xy)^{\frac{1}{2}}], \\ &(\partial/\partial y)\varphi(x, y) = \varphi(x, y) - (\partial/\partial y) I_0[2(xy)^{\frac{1}{2}}], \\ &\varphi(0, y) = 0, \quad \varphi(x, 0) = e^{x} - 1. \end{aligned}\right\} \qquad (17)$$

Tables of a related function,

$$g(x, y) = e^{-x^2-y^2} \varphi(x^2, y^2), \qquad (18)$$

have been constructed[9] and are to be published. By means of Eqs. (17), Eq. (15) is readily shown to satisfy Eqs. (12) and (13), and this solution is unique.[6]

An expression for the quantity Γ can be obtained by the integration of either of Eqs. (11), employing Eqs. (15) and (17). Expressions for the temperatures of solid and fluid can then be obtained by the combination of this result with the solution for the variable $\Delta$, Eq. (15). However, the temperature of the fluid is obtained more directly from the relation,

$$\partial V/\partial \xi = \Delta e^{-\xi-(1-\beta)\tau}, \qquad (19)$$

which results from the substitution of the definition of $\Delta$ in the second of Eqs. (6). The temperature of the solid is then obtained at once by the combination of this result with the expression for $\Delta$.

## THE TEMPERATURES OF SOLID AND FLUID

If the solution for $\Delta$ be combined with Eq. (19), there is obtained

$$\frac{\partial V}{\partial \xi} = \left\{ U_0 I_0[2(\xi\tau)^{\frac{1}{2}}] + \frac{\alpha}{1-\beta} \varphi[\tau(1-\beta), \xi/(1-\beta)] \right\} e^{-\xi-(1-\beta)\tau}. \qquad (20)$$

Integrating this expression between limits as an ordinary differential equation, and recalling that $V(0, \tau) = 0$, one obtains

$$V = U_0 \varphi(\xi, \tau) e^{-\xi-(1-\beta)\tau} + \frac{\alpha}{1-\beta} e^{(\beta-1)\tau} \int_0^{\xi} e^{\xi'} \varphi[\tau(1-\beta), \xi'/(1-\beta)] d\xi'. \qquad (21)$$

The integral of Eq. (21) can be evaluated by the method of integration by parts, employing the properties of $\varphi(x, y)$ which are summarized by Eqs. (17). The result of this process is

$$\int_0^{\xi} e^{-\xi'} \varphi[\tau(1-\beta), \xi'/(1-\beta)] d\xi' = \frac{1-\beta}{\beta} \{ \varphi(\xi, \tau) - \varphi[\xi/(1-\beta), \tau(1-\beta)] + [e^{\xi/(1-\beta)} - e^{\xi}] e^{\tau(1-\beta)} \} e^{-\xi}, \qquad (22)$$

if $\beta \neq 0$, and

$$\int_0^{\xi} e^{-\xi'} \varphi[\tau(1-\beta), \xi'/(1-\beta)] d\xi' = \{ \xi\varphi(\tau, \xi) + \tau\varphi(\xi, \tau) - (\xi\tau)^{\frac{1}{2}} I_1[2(\xi\tau)^{\frac{1}{2}}] \} e^{-\xi}, \qquad (23)$$

---

[7] H. Bateman, *Partial Differential Equations of Mathematical Physics* (The Cambridge University Press, New York 1932), p. 128.

[8] H. C. Thomas, J. Am. Chem. Soc. 66, 1664 (1944).

[9] S. R. Brinkley, Jr. and R. F. Brinkley, unpublished tables.

if $\beta=0$. $I_1$ is the first-order Bessel function of the first kind for purely imaginary argument. Therefore, the temperature of the fluid is given by

$$V(\xi,\tau)=\left(U_0+\frac{\alpha}{\beta}\right)\varphi(\xi,\tau)e^{-\xi-(1-\beta)\tau}+\frac{\alpha}{\beta}\{e^{\beta\xi/(1-\beta)}-\varphi[\xi/(1-\beta),\tau(1-\beta)]e^{-\xi-(1-\beta)\tau}-1\},\tag{24}$$

if $\beta\neq0$, and by

$$V(\xi,\tau)=\{(U_0+\alpha\tau)\varphi(\xi,\tau)+\alpha\xi\varphi(\tau,\xi)-\alpha(\xi\tau)^{\frac{1}{2}}I_1[2(\xi\tau)^{\frac{1}{2}}]\}e^{-\xi-\tau},\tag{25}$$

if $\beta=0$. The temperature of the solid is immediately obtained by combination of these results with the expression for $\Delta$. Then,

$$U(\xi,\tau)=\left(U_0+\frac{\alpha}{\beta}\right)\{e^{\beta\tau}-\varphi(\tau,\xi)e^{-\xi\ (1-\beta)\tau}\}-\frac{\alpha}{\beta}\left\{\frac{1}{1-\beta}\varphi[\tau(1-\beta),\xi/(1-\beta)]e^{-\xi-(1-\beta)\tau}+1\right\},\tag{26}$$

if $\beta\neq0$, and

$$U(\xi,\tau)-U_0=\{[\alpha(\xi+1)-U_0]\varphi(\tau,\xi)+2\tau\varphi(\xi,\tau)-\alpha(\xi\tau)^{\frac{1}{2}}I_1[2(\xi\tau)^{\frac{1}{2}}]\}e^{-\xi-\tau},\tag{27}$$

if $\beta=0$. The rate of heat transfer from solid to fluid per unit volume of porous solid is equal to $h\Delta(\xi,\tau)c^{-\xi-(1-\beta)\tau}$.

The application of these results to particular systems will be the subject of subsequent reports from this laboratory.

## THE STEADY STATE

The steady state values of the temperatures of fluid and solid are obtained from Eqs. (24) to (27) by proceeding to the limit of infinite $\tau$. The function $\varphi(\xi,\infty)$ is finite for finite $\xi$, and the function $\varphi(\tau,\xi)$ tends to $e^{\xi+\tau}$ as $\tau$ tends to infinity. Then, one obtains

$$V(\xi,\infty)=(\alpha/\beta)\{e^{\beta\xi/(1-\beta)}-1\},\quad U(\xi,\infty)=\frac{\alpha}{\beta}\left\{\frac{1}{1-\beta}e^{\beta\xi/(1-\beta)}-1\right.\tag{28}$$

if $\beta\neq0$, and

$$V(\xi,\infty)=\alpha\xi,\quad U(\xi,\infty)=\alpha(\xi+1),\tag{29}$$

if $\beta=0$. These steady state solutions have been previously obtained and employed for the calculation of catalyst and gas temperatures in a typical operation of the Michael process for the synthesis of hydrocarbons from carbon monoxide and hydrogen.[10]

---

[10] S. R. Brinkley, Jr., 110th Meeting, American Chemical Society, Chicago, Illinois, Sept. 1946.

# Letters to the Editor

## Concerning Estimates of the Minimum Sparking Potential Based upon the Cathode Work Function

DONALD H. HALE AND W. S. HUXFORD
*Department of Physics, Northwestern University, Evanston, Illinois*
April 4, 1947

JACOBS and LaRocque[1] have recently suggested a method for estimating the minimum sparking potential for cold cathode tubes. They measured the minimum sparking potentials for several cathodes in argon and calculated the corresponding value for $\gamma_m$. These values for $\gamma_m$ were then plotted against the corresponding values of the work functions and the minimum sparking potentials. By use of these two curves an attempt is made to estimate $\gamma_m$ and the minimum sparking potential in argon for any cathode, provided only the work function for the cathode material is known. Applying this method the authors obtain an approximate and reasonable value for the minimum sparking potential of a sodium cathode in argon.

This method of determining $\gamma_m$ from sparking data is not reliable. It has been pointed out that this procedure cannot be expected to yield good values for $\gamma$ in all cases.[2] This is shown by the results obtained by these authors since their values of $\gamma$ vary widely from those found by other workers. Schöfer[3] and Engstrom[4] found that the values of $\gamma$ for a barium cathode in argon increased with increasing values of $E/p$ starting with an $E/p$ of about 40. The authors found just the opposite. Their values of $\gamma$ decrease with increasing values of $E/p$. Further, their value of $\gamma$ at an $E/p$ of about 40 is three times greater than that measured by Schöfer. Since they worked at higher gas pressures than did earlier workers it may be that in their work the mechanism of electron emission by photons at the cathode was enhanced. However it is difficult to see how this could account entirely for their high values of $\gamma$.

The agreement between the calculated and experimental values of the minimum sparking potential for a sodium cathode in argon may be fortuitous. Bowls[5] measured $\gamma$ for a sodium cathode in nitrogen and Ehrenkrantz[6] measured the sparking potentials. Bowls found that the sodium cathode gave distinctly lower values of $\gamma$, for a wide range of values of $E/p$, than did a platinum cathode. Yet, Ehrenkrantz found that the sodium cathode showed a lower minimum sparking potential than did the platinum. Hale[7] and Ehrenkrantz found a similar situation with regard to sodium cathodes in hydrogen. The authors explain the lower $\gamma$ for sodium on the grounds that hydrogen or nitrogen combines with sodium to give a surface of high work function. Such an assumption would explain the low values for $\gamma$ but it cannot explain, on the basis of the author's argument, the observed low value of the minimum sparking potential. Actually it has been pointed out that sodium surfaces in hydrogen or nitrogen show low photoelectric thresholds.[8] Here then is a case of a cathode material which has a low work function, which shows low values of $\gamma$, in at least two gases, and yet has a low minimum sparking potential.

There are other interesting examples. The work functions for nickel and platinum are 5.01 and 6.30 volts, respectively.[9] Yet the minimum sparking potential for argon with cathodes of these materials is 195 volts in each case.[10] The work function for iron is 4.72 volts.[9] However, the minimum sparking potential for argon with an iron cathode is 265 volts. This is distinctly higher than the minimum sparking potential for platinum in argon. Yet platinum has the higher work function.

It is probably correct to say that, in a very general way, materials with low work functions will show high values of $\gamma$. However, exceptions are known and any prediction of the minimum sparking potential which is based upon the work function of the cathode may be badly in error. Much work remains to be done before this point is clear.

We wish to call attention to what appears to be a misstatement. The authors state that Bowls found a lower $\gamma$ for platinum in a large region of $E/p$ than he did for sodium. It appears that they intended to refer to a lower $\gamma$ for sodium which was what Bowls observed.

[1] H. Jacobs and A. P. LaRocque, J. App. Phys. **18**, 199 (1947).
[2] L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases*, pp. 348 and 416.
[3] R. Schöfer, Zeits. f. Physik **110**, 21 (1938).
[4] R. W. Engstrom, Phys. Rev. **55**, 239 (1939).
[5] W. E. Bowls, Phys. Rev. **55**, 293 (1938).
[6] F. Ehrenkrantz, Phys. Rev. **55**, 219 (1939).
[7] D. H. Hale, Phys. Rev. **55**, 815 (1939).
[8] L. B. Loeb, reference 2, p. 395.
[9] A. L. Hughes and L. A. DuBridge, *Photoelectric Phenomena*, p. 75.
[10] M. J. Druvesteyn and F. M. Penning, Rev. Mod. Phys. **12**, 114 (1940).

---

## Measurement of the Dynamic Stretch-Modulus and Hysteresis of Tire Cords

W. JAMES LYONS AND IRVEN B. PRETTYMAN
*The Firestone Tire and Rubber Company, Akron 17, Ohio*
April 17, 1947

THE modulus of elasticity (Young's) of synthetic filaments and films has been measured by Meyer and Lotmar[1] and Ballou and Silverman,[2] who employed sound-velocity methods which gave the dynamic value of the modulus at acoustical frequencies. These methods, however, are not applicable to the measurement of concomitant hysteretic losses. Wakeham and Honold[3] studied the elastic modulus and hysteresis effects in cotton, rayon, and Nylon tire cords under dynamic conditions, but at a relatively low frequency: 1 cycle/sec.

The exploratory studies reported herein were undertaken to develop a method for measuring not only the dynamic

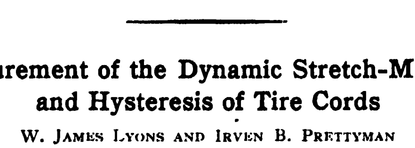


FIG. 1. Schematic diagram of cord stretch-vibrometer.

TABLE I. Representative measurements of dynamic properties of conventional tire cords, obtained on cord vibromete at constant impressed force-amplitude.

| Cord sample | | Static properties (conditioned at 65% R.H., 70°F) | | | Dynamic properties (at approximately 14% R.H., 78°F) | | |
|---|---|---|---|---|---|---|---|
| Fiber | Construction | Breaking load (Kg) | Ultimate elongation (Percent) | Resonant frequency, $f_0$ (cycles/sec.) | Dynamic modulus, $E$ ($10^{10}$ dynes/cm$^2$) | Internal friction, $\eta$ ($10^6$ poises) | Hysteresis index, $2\pi f\eta$ ($10^{10}$ dynes/cm$^2$) |
| Cotton | 11/4/2 | 8.5 | 14.3 | 169 | 7.7 | 4.0 | 0.42 |
| Viscose rayon | 1100/2 | 6.8 | 15.9 | 148 | 11.2 | 5.3 | 0.49 |
| Nylon | 210/3/3 | 14.6 | 19.5 | 116 | 6.1 | 2.4 | 0.17 |

modulus of filaments, yarns, and cords in the sonic or near-sonic range, but also their internal frictional (hysteretic) properties. Measurements of the modulus and friction have been successfully made on cotton, rayon, and Nylon cords forcibly vibrated longitudinally at frequencies above 100 cycles/sec., by electromagnetic means.

The apparatus is shown schematically in Fig. 1. A small light-weight coil is placed coaxial with the core of a strong electromagnetic field coil, so that the former is in a powerful magnetic field when a direct current passes through the field coil. The small coil is rigidly mounted on a plastic rod which passes through a hole at the axis of the core. The rod is supported by pieces of the test cord attached coaxially at each end. Static loads at the outer ends of the two pieces of cord balance each other and allow the cords to be placed under tension without displacing the rod and small coil. When an oscillating current is sent through this coil, the rod vibrates longitudinally, causing small periodic extensions and contractions in the pieces of test cord. In the initial trials, a small mirror was rotatably connected to the rod at the end opposite the vibrating coil, so that by means of an optical-lever system, amplitude of vibration could be detected and measured on a ground-glass scale. It was subsequently found that a micrometer microscope could be advantageously substituted for the optical-lever system, with improved accuracy; and this method was used to obtain the results reported herein. In the present experiments the vibrating coil was energized by a 50-watt power amplifier which was controlled by a regular commerical calibrated audiofrequency oscillator unit.

The principle underlying the present method is expressed by the equation of motion of the vibrating element under impressed force:

$$m(d^2s/dt^2)+(\eta/q)(ds/dt)+(E/q)s=F\cos\omega t, \qquad (1)$$

where $m$ = vibrating mass, $s$ = displacement from the equilibrium position; $\eta$ = coefficient of internal friction (hysteresis); $E$ = dynamic stretch modulus; $q$ = shape factor of cord; and $F$ = amplitude of applied force having frequency $\omega/2\pi$. Equation (1) has been applied to forced compressive vibrations in flat cylinders of rubber by Gehman, Woodford, and Stambaugh,[4] and to shear vibrations in similar specimens by Dillon, Prettyman, and Hall.[5] The design of the driving mechanisms used by these groups provided the prototype for the corresponding elements of the present equipment.

It can be shown readily that at resonance, to good approximation:

$$E=mq\omega_0^2=m(L/2A)(2\pi f_0)^2, \qquad (2)$$

where $A$ = cross-sectional area of cord or yarn (exclusive of interstices between fibers); $L$ = free length of each matched piece of cord, and $f_0$ = resonant frequency.

Equation (2) has been applied to observations of resonant frequencies made with the cord stretch-vibrometer on three kinds of tire cord, with the results summarized in Table I. To identify the cords the conventional static properties have been included in the table. All cord samples were vibrated under a tensional load of 1.2 kg, with $L=25.4$ cm (10 in.). The effective mass of the vibrating system, $m$, was found to be 17.0 g. The cross-sectional area of each sample was determined from the weight of a 200-cm specimen of the cord and the bulk density of the component fiber.

Internal friction, $\eta$, may be readily derived from observations on the stretch-vibrometer by means of the equation:

$$\eta=Fq/s_m\omega_0, \qquad (3)$$

$s_m$ being the maximum (resonance) strain-amplitude. Values of this quantity, as well as of an index of hysteresis losses $2\pi f\eta$, are entered in Table I. The latter quantity has been shown, in the case of rubber,[5] to be substantially independent of frequency.

---

[1] K. H. Meyer and W. Lotmar, Helv. Chim. Acta **19**, 68 (1936).

[2] J. W. Ballou and S. Silverman, Textile Research **14**, 282 (1944), J. Acoust. Soc. Am. **16**, 113 (1944).

[3] H. Wakeham and E. Honold, J. App. Phys. **17**, 698 (1946).

[4] S. D. Gehman, D. E. Woodford, and R. B. Stambaugh, Ind. Eng. Chem. **33**, 1032 (1941).

[5] J. H. Dillon, I. B. Prettyman, and G. L. Hall, J. App. Phys. **15**, 309 (1944).

---

## On the Analysis of Internal Flow Machinery

SABINUS H. CHRISTENSEN
*Carrier Corporation, Syracuse, New York*

April 11, 1947

THIS is a brief report of a new method of applied research which has opened an area of analysis hitherto inaccessible to internal flow machinery having moving passages (centrifugal and axial-flow compressors, and turbines).

The new technique is based on the rejection of the idealization in the gas-dynamic analysis of internal flow machinery that the fluid may be considered to be a compressible ideal gas, i.e. frictionless. It has been found by exhaustive analytic study of test data that the use of the frictionless gas concept fails to explain performance within very wide limits.

It is well known that the inefficiency of internal flow machinery is caused by the dissipation of available energy which is reflected in increase of entropy through the machine. The basic domain for research on internal flow machinery is then recognized to be that of making a detailed explanation of entropy increase caused by friction, separation, and turbulence. Taking the detailed change of entropy into full account now explains such hitherto inaccessible phenomena as quantitative values of shock loss at the entrance to rotating grids, and the theory of the maximum flow in a centrifugal compressor (as have been done by the writer in unpublished company reports).

The new method is directed towards the local evaluation of entropy in rapidly moving or rotating passages. The method utilizes the device of speeding up the moving passage (or the equally effective method of using a very heavy gas with slower passage speeds) until somewhere, within the passage the relative velocity is locally sonic. The exact equations of state, energy, and continuity are then satisfied by a definite entropy value at the blocking point. If a study is made of the energy *vs.* area gradient along the passage it will be seen that the blocking point is fixed by the passage design, thus providing the fourth simultaneous equation which solves the problem of determining the exact thermodynamic state inside a rapidly moving passage. It is believed that this technique is far superior to the less reliable method of trying to measure the state of the fluid by instruments which are forced to transmit their data through commutator rings and packings.

Instead of working with entropy change, re-arrangement of the gas-dynamic equations shows that the isentropic relations apply if the passage area is modified by the factor $\exp(-J\Delta S/R)$, $J$ being Joule's mechanical equivalent of heat, $\Delta S$ the increase in entropy, and $R$ the gas constant.

---

## Objective Aperture System for the Electron Microscope

CECIL E. HALL
*Massachusetts Institute of Technology, Cambridge, Massachusetts*

April 17, 1947

THE conventional objective aperture used in magnetic electron microscopes consists of a circular diaphragm about 25 microns in diameter, fixed with respect to the lens and coaxial with it. It prevents electrons scattered through angles larger than about $10^{-2}$ radian from reaching the image plane, thereby reducing background intensity and enhancing contrast in the image. An objective aperture is necessary for the examination of thick objects producing a high scattered intensity, such as replicas, sections, and the like, but it is of lesser value in the examination of specimens of suitable thinness and contrast. Electron microscopists are familiar with the many practical difficulties associated with the use of apertures. They are difficult to make, install, and align, and must be replaced frequently, because they accumulate films of non-conducting material which become charged and produce distortions in the image. It is therefore desirable to have an objective aperture system that can be installed quickly and renewed as occasion demands.

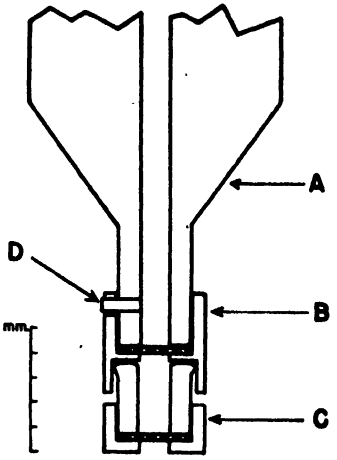


FIG. 1. Objective aperture system.

An objective aperture system having the advantage that it is attached to the specimen cartridge rather than to the lens, is shown in Fig. 1. It has been used successfully in this laboratory with an RCA type B electron microscope. The device consists essentially of a second screen mounted and aligned so that each opening of the specimen screen has below it a corresponding opening serving as an aperture stop. Aperture screens may be detached readily, if they are not required for the specimen under examination, and may be renewed frequently from standard material. *A* is the lower part of the specimen cartridge to which is fitted the cap, *B*, containing the specimen. The tip, *C*, containing the aperture screen is mounted in a friction cup so that it may be turned about the axis and rocked slightly to give small lateral displacements. Aperture screens, which must be non-magnetic since they are in the lens field, are punched from 200-per-inch copper Lektromesh.[1]

The procedure in using the device is as follows: Clip the cartridge to the stage of a 100-power optical microscope with cross-hairs in the eyepiece. Focus on the object screen and rotate the eyepiece until the cross-hairs are aligned with the screen wires. Focus on the aperture screen and rotate the tip *C* until the screen wires are square with the cross-hairs. Refocus on the object and center an open square on the cross-hairs with the mechanical stage. Focus on the aperture mesh and rock the tip, *C*, until an open square is centered on the cross-hair intersection. This procedure usually takes one or two minutes. Possibly a smaller object-to-aperture distance than the three milli-

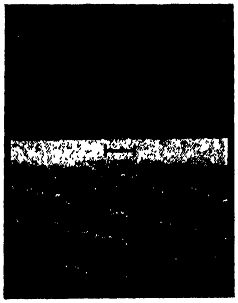


FIG. 2. Collodion-silicon replica of glass grating, shadowed with chromium; (a) without an aperture, (b) with mesh aperture. X3600.

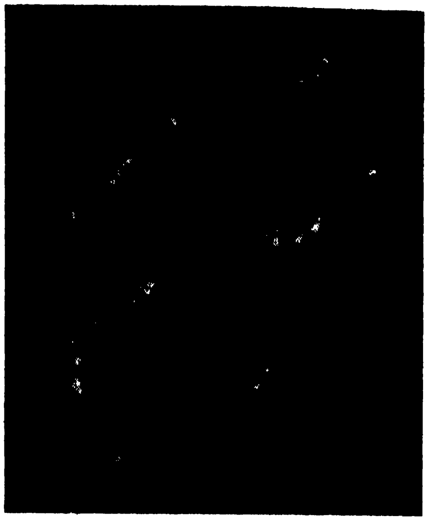


FIG. 3. Dark-field electron micrograph of carbon particles shadowed with uranium. ×11,666.

meters shown in the diagram would provide adequate contrast in the image and at the same time simplify the alignment of the two meshes.

Two micrographs of a collodion-silicon replica of a glass grating are shown in Fig. 2, to demonstrate the difference in contrast with and without the aperture screen; (a) was taken without an aperture, while (b) was taken with one. The replica was shadowed with chromium. Contacts were made from the original micrographs on the same Eastman Medium lantern slide plate and a contact print made from this on F-3 Kodabromide. Contrast, of course, depends on the photographic processing, but an effort was made to have the prints represent a reasonable comparison of the original plates. Although improved contrast in the image is most pronounced with thick specimens, improvement has also been noted in thin specimens containing small particles of low scattering power. Open areas in standard Lektromesh screens are square, but no distortions have been noted in the image which could be attributed to this unusual shape. Also, it is possible that electrons may be so widely scattered that they will reach the image plane through an opening other than the one directly below the point under observation. In the mesh normally employed the solid parts are three times as wide as the open squares, and since scattered intensity falls off rapidly with angle, the proportion of such widely scattered electrons appears to be inappreciable.

The multiple aperture was originally devised to enhance contrast in images of thick specimens, but it has proved to be very useful also for dark-field electron microscopy. If alignment of the two meshes is not perfect, and it very rarely is, a part of each object opening will be seen in dark field. An example of a dark-field micrograph made in this manner is reproduced in Fig. 3. The specimen consisted of carbon particles[2] on a collodion film, shadowed with uranium to a calculated thickness of about 13A. Shadow areas are dark because they contain no uranium, hence scatter least. The uranium film appears as a background of discrete spots ranging in diameter from 50 to 200A. The spots are quite well resolved for dark field, indicating that they are produced by electrons that have been scattered coherently and are therefore confined to a narrow angular aperture and velocity range. The dark-field micrograph thus reveals the presence of small ordered regions in the film that are not perceptible as such in bright field. Similar structure has also been observed with thin films of several other metals and compounds.

Images of the carbon particles shown in Fig. 3 are characterized by bright contours and dark centers, resulting from the fact that at small angles there is a relatively high intensity from electrons scattered close to the surface. Electrons scattered within the particles are spread over a much larger solid angle so that a smaller proportion enter the limiting aperture. An interesting feature, not visible in bright field, is the occurrence of numerous bright spots at the edges of the particles. This appearance is not associated with the uranium since the spots also appear in unshadowed specimens. In the micrograph shown, all such spots appear on edges *away* from the aperture, hence represent pencils that have been deflected *under* the particles. However, the occurrence on one edge only, is dependent on alignment of the electron microscope and on the objective current. They have been observed on edges toward or away from the aperture and on all sides simultaneously. The reason for enhanced scattered intensities is not evident at present. Possibly they are in the nature of crystalline reflections owing to regularities near the surface, although the images are considerably more diffuse than those produced by diffractions from common crystalline substances. Whether the enhanced scattering is to be accounted for on the basis of crystallinity or not, it must be noted that near the edges, electrons are incident and emergent at such small angles that refraction effects may be expected to be of significant magnitude. The observed deflection of electrons under the particles could be accounted for on the basis of refraction resulting from an inner potential of the order of 10 volts.

These observations demonstrate the utility of the multiple aperture which is, to be sure, a compromise for convenience rather than the equivalent of a clean, well-centered aperture of the conventional type. Probably a completely satisfactory solution of the aperture problem will require the redesign of present objectives to allow for insertion and alignment of apertures fixed to the lens, without the difficulties and hazards that presently attend this operation. As an alternative, the multiple aperture has some advantages, and furthermore, the ease with which it may be used for dark-field observations may stimulate this interesting and somewhat neglected aspect of electron microscopy.

---

[1] C. O. Jelliff Manufacturing Corporation, Southport, Connecticut.
[2] P33, Thermatomic Carbon Company, Inc.

# New Books

## A Shorter History of Science

By Sir William Cecil Dampier. The Macmillan Company, New York, 1944.

When Sir William Cecil Dampier's larger book, *A History of Science and Its Relations with Philosophy and Religion*, came from the press, we recognized that a great service had been rendered: science had been given its "setting in other modes of thought." Could Dampier streamline the larger book and produce *A Shorter History of Science* without devitalizing the story of the growth of science which made the larger book famous? In the condensation, the story has been more simply told, but even so its vitality has been preserved. The connections of science with other human activities have been reduced to the more direct impacts, but these fewer situations have been used expertly to show the relations of science to philosophy and religion. Undoubtedly the shorter history, like its monumental prototype, will become known, not simply for the facts and dates which it contains, but especially for the story it tells of the expanding mind and spirit of man.

"Science has two streams," we read in the first chapter, "corresponding to two sources, the first a gradual invention of tools and implements whereby men earn their living more safely and easily, and second the beliefs they form to explain the wonderful universe around them. The first may perhaps better be called technology, for its problems are too difficult for early theory, and only in later stages does it become applied science; the second, which in historic times grew into a pure search for knowledge, is the main subject of this book."

As one reads the book's eleven chapters—with titles such as: the origins, Greece and Rome, Galileo and Newton, the eighteenth century, recent biology, the new physics and chemistry, and the stellar universe—one turns back to the Preface for this summary: "The Greek Atomists, besides speculating on the structure of matter, developed therefrom a mechanical theory of life. Conversely, the philosophy of Plato, as modified by Aristotle, laid too much stress on innate ideas and logical deduction to make a favorable background for the beginnings of inductive experimental science. To Newton and his immediate followers the Heavens declared the Glory of God, but Newton's work produced a very different effect on the minds of Voltaire and other eighteenth-century sceptics. Darwin's revival of the old theory of evolution on the new basis of natural selection not only suggested an alternative origin for mankind, but spread evolutionary doctrine far beyond the limits of biology. The recent revolution in physics has shaken the evidence for philosophic determinism which the older synthesis seemed to require. Such broad effects must be noticed, but the more technical aspects of philosophy may be passed by."

Exponents of general education, who insist that all courses should be integrated into the whole life pattern of the learner, will like the spirit of *A Shorter History of Science*. Its author thinks that "early specialization is dangerous," and writes, "Those older schoolboys whose chief subjects are scientific should look at them also from a humanist standpoint, and realize their setting in other modes of thought, while those studying literature need some knowledge of science before they can be said to be well educated. For both groups, I believe that the history of science, the story of man's attempts to understand the mysterious world in which he finds himself, makes the best way of approach to common ground."

Carl F. Eyring
*Brigham Young University, Provo, Utah*

## Applied Atomic Power

By E. Smith, A. Fox, R. Sawyer, and H. Austin. Pp. 227, 6¼×9¼ in. Prentice-Hall, New York, 1946. Price $4.00.

This book with a rather ambitious title is easy and interesting reading uncomplicated by mathematics. Undergraduates will find everything simple except perhaps the descriptions of nuclear physics work at Los Alamos in Appendix I, which had best be read after reading Part III.

Part I contains a description of 32 uranium-bearing minerals and a history of radioactivity. Part II is introductory physics, mostly off-the-slide-rule stuff. Part III is an abstract of the Smyth Report with two noteworthy portraits of Fermi and Dunning, and a description of the three production piles at Hanford. The description of the Clinton Engineer Works is more sketchy because of security reasons.

Part IV, comprising the meat of the book, describes in considerable detail possible methods of using atomic power. The author, Sawyer, projects a 35,000 hp AK gas turbine using a closed cycle with a helium-carbon dioxide mixture. The inert helium makes possible a simple uranium heater. Many arrangements are discussed using fission heat in place of one fuel or another. It is emphasized that atomic energy will never replace other fuels but will supplement them. For some uses it is suggested that water be used as a moderator-coolant which is vaporized and then used to run a turbine. Atomic energy locomotives, ships, and rockets are considered.

Industrial benefits from gaseous-diffusion research are the subject of Part V.

Appendix III lists nuclear and atomic masses of isotopes from hydrogen to gold-197. A seven-page index is provided.

Duane Studley
*Colorado Springs, Colorado*

## Analytical Experimental Physics (Revised Edition)

By Harvey B. Lemon and Michael Ference, Jr., University of Chicago. Pp. 588+XVI, Figs. 584, Plates 66, 24×30 cm. University of Chicago Press, 1946. Price $8.00.

This college textbook first published in 1942 was hailed as the "Colossus of General Physics, one of the seven wonders of the textbook world." Now it has been revised, corrections have been made, rough spots have been smoothed out, figures, plates, and pages have been added. It is indeed a colossus in quantity and quality.

Little revision has been made in the chapters on mechanics, while in the section on heat qualitative discussions of the quantum theory of specific heats and of the properties of dilute solutions have been added. The chapter on alternating currents now includes a section on parallel circuits containing resistance, inductance, and capacitance. Much has been added to the chapter on conduction through gases so that it now includes a discussion of the motion of charged particles in electric and magnetic fields, of relativistic variation of mass with velocity, of the cyclotron, and of the betatron. As is to be expected the sections on nucleonics and electronics have been greatly expanded. A discussion is given of some nuclear reactions of tracers, nuclear fission, transuranic elements, nuclear chain reactions, and the atomic bomb. The electronic chapter now includes an analysis of the equivalent circuit of a triode, a discussion of gas-filled tubes, pentodes, photoelectric cells, electronic oscillators, radio transmitting and receiving apparatus, microelectric waves with radar transmitting and receiving systems, television, and the electron microscope. An interesting plate is added on radar, showing typical sending and receiving apparatus, a "P.P.I. scope" presentation of a convoy leaving New York harbor, and the trace in an oscilloscope of the reflection of radar waves from the moon. Some additions have been made to the chapter on wave motion and only minor changes in the chapters on sound and light. A useful set of trigonometric and logarithmic tables, together with a table of important physical constants, has been appended.

The book is what its title implies, both analytical and experimental in its approach to physics. Physical concepts are carefully interwoven with experimental data. The mathematics employed in the analysis is largely algebra and trigonometry with calculus used sparingly. The presentation is excellent and of the kind to appeal to the abler students. Though the book is more inclusive and more difficult than almost any other college one-year general text in physics, the authors have starred about one hundred sections which may be omitted so as to bring it within the scope of most college courses. There is, however, a section on the equivalent circuit of a triode which would appear to be beyond even the abler students with their present knowledge of physics. The topics are discussed in such a manner as to leave the student with a clear idea of the fundamental concepts and how these were developed.

Since the book is so large and heavy it might have been more practical to publish it in two smaller volumes. Whether this be good or bad advice the authors are to be congratulated on the challenge which they have given to students. Any student who masters this material can be sure of having a solid foundation in physics.

R. J. Stephenson
*The College of Wooster*

# New Booklets

Allied Radio Corporation, 833 West Jackson Boulevard, Chicago 7, Illinois, has announced the publication of a new 164-page catalog for 1947, covering radio and electronic products. Special emphasis is placed on equipment for industrial maintenance, research, and production requirements as well as for the needs of government agencies. 10,000 items. Free on request.

The Superpressure Division of the American Instrument Company, Silver Spring, Maryland, has issued its new 88-page catalog No. 406, entitled *Superpressure Catalytic Hydrogenation Apparatus.* Inquire on company letterhead.

United States Testing Company, Inc., 1415 Park Avenue, Hoboken, New Jersey, has revised its booklet *Testing Terms for Better Understanding.* The book describes some of the important tests of merchandise—tells how tests are made, what the results mean, what are standard and minimum specifications.

The Scientific Personnel Branch of the Navy Department, Washington 25, D. C., issues bi-monthly a *Scientific Personnel Bulletin*, listing employment opportunities in the Navy Department for scientists and technicians. 76 pages in the January-February, 1947, issue.

Eberbach and Son Company, Ann Arbor, Michigan, featured in its March issue of *The Announcer of Scientific Equipment* an article on the background and development of the modern analytical balance. 16 pages. Available on request.

Fischer and Porter Company, Hatboro, Pennsylvania, has released a new bulletin including dimensional and operational details of an improved continuous liquid specific gravity indicator. Address Department 6F-B of the company.

National Carbon Company, Inc., 30 East 42 Street, New York 17, New York, has published a new bulletin, *Catalogue Section M8808*, describing standard, seven-tube heat exchangers of "Karbate" impervious graphite for use under highly corrosive conditions.

The Winter 1946–47 number of *Interchemical Review*, published at 432 West 45th Street, New York 19, New York, has the following table of contents:

Maleic and Fumaric Resins....Charles S. Rowland.
Market Research in Industrial Fields....John P. Duane.
Oilcloth Pioneers in Maine....Charles R. Bragdon.

The Gaertner Scientific Corporation, 1201 Wrightwood Avenue, Chicago 14, Illinois, has recently published a new bulletin, 156-74, describing its optical benches and accessories. 26 pages.

# Here and There

### New Appointments

Eugene W. Boehne and Lan Jen Chu were recently appointed associate professors in the Department of Electrical Engineering at the Massachusetts Institute of Technology.

Two new appointments to the staff of the National Bureau of Standards were recently announced. Albert S. Cahn, Jr., will assist in the research involved in the design of two high speed electronic computing machines to be built for the Bureau of the Census and the Navy. P. J. Selgin will work on the development of electronic ordnance for the military services in the Ordnance Development Division.

The U. S. Weather Bureau has announced the appointment of Ross Gunn as the director of a newly organized Division of Physical Research, which will undertake studies of the atmosphere and basic physical processes that are related to the weather and the earth.

F. C. Henriques, Jr., has joined the staff of Tracerlab, Inc., at Boston, Massachusetts, as director of the Radiochemical Division

### Awards

Latest Franklin Institute awards for 1947 are as follows:

The Franklin Medal, highest honor of the Institute, to Enrico Fermi, physicist at the Nuclear Research Institute in Chicago, and to Sir Robert Robinson, professor of chemistry at Oxford University, England.

The Newcomen Medal to Everett G. Ackart, recently retired chief engineer of the E. I. du Pont de Nemours and Company.

The Clark Medal to Edward G. Boyer, manager of the Philadelphia Electric Company's gas department.

The Ballantine Medal to George C. Southworth of the Bell Telephone Laboratories.

### New Chairman for A. I. P.

George R. Harrison, dean of the School of Science, Massachusetts Institute of Technology, has been elected chairman of the American Institute of Physics, succeeding Paul E. Klopsteg, director of research, Northwestern University Technological Institute, who has served as chairman since 1940.

### Western Branch for Airborne Instruments Laboratory

A branch of the Airborne Instruments Laboratory, Inc., Mineola, New York, was recently established at Burbank, California. This will make possible a closer liaison between the main laboratory and west coast aircraft manufacturers in need of a consultation and development service. Airborne Instruments Laboratory is the peacetime outgrowth of the NDRC laboratory of the same name which operated under the direction of Columbia University during the war. Hector R. Skifter is president of the Laboratory and Robert D. Martin is supervising the west coast unit.

### Electrochemical Symposium at Pittsburgh

The annual spring symposium of the Pittsburgh section of the Electrochemical Society was held May 23 and 24 at the Mellon Institute. The subject was "High Temperature and Hetrogeneous Reactions."

### Electrochemical Congress at Louisville

A congress sponsored by the Electrochemical Society met in Louisville, Kentucky, April 9 to 12. A day's session was devoted to electronics, one to the corrosion of metals at elevated temperatures, and a third to the production of the best steels in the electric furnace. The Young Authors'. Prizes were awarded to N. A. Neilsen of Du Pont and to Burke Cartwright of the U. S. Bureau of Mines.

### Engineering Foundation Issues Report

The Engineering Foundation, 29 West 39th Street, New York 18, New York, made public in March its 32nd annual report to the board of trustees. The report stated, "It is clear that industry has, in general, become more research minded and large expansions of industrial research are to be expected. In some cases this will be by individual companies, in others, by associations of various types." Sixteen projects were sponsored by the Engineering Foundation during the past year, six of which were carried over from the preceding year. Among the projects were soil mechanics, fluid mechanics, materials of engineering comprising metal researches, welding, lubrication, and rolling friction. In addition to sponsoring engineering research, the Foundation also supports agencies or activities having as their objectives the advancement of the engineering profession.

### Changes in Electrical and Photometric Units

In pursuance of decisions of the International Committee on Weights and Measures, the National Bureau of Standards will introduce as of January 1, 1948, revised values of the units of electricity and of light. While the definitions of the units and the methods of fixing their magnitudes will be different from the present practical systems, the changes in magnitude will be so small as to affect appreciably only measurements of high precision. In certificates for standards and instruments issued by the Bureau during 1947, values will be given in both the old and the new units.

A complete description of the new units has appeared in the *Review of Scientific Instruments* for May, 1947.

Journal
of
Applied Physics

Volume 18, Number 7 July, 1947

# The Nucleation of Ice Formation by Silver Iodide

B. VONNEGUT
*General Electric Research Laboratory, Schenectady, New York*

(Received March 17, 1947)

Silver iodide particles have been found to serve as nuclei for the formation of ice crystals in super-cooled water and in water vapor super-saturated with respect to ice. It is believed that silver iodide serves as a very effective nucleus because it very closely resembles ice in crystal structure. Both dimensions of the unit cell of ice and silver iodide are the same to within approximately one percent. The maximum temperature at which the silver iodide particles serve as nuclei is approximately −4°C for particles one micron in diameter, and −8°C for particles 100 Angstrom units in diameter. A silver iodide smoke generator has been constructed which consumes 1 mg of silver iodide per second and produces $10^{15}$ effective nuclei per second.

V. J. SCHAEFER of this laboratory has recently reported that super-cooled liquid water clouds spontaneously transform from water to ice when the temperature is −35°C or lower. He observed that at temperatures as low as −20°C, the addition of fine particles of a wide variety of substances was without effect on the super-cooled cloud.[1]

B. M. Cwilong,[2] using a Wilson cloud chamber, found that moisture in clean air is precipitated as a cloud of ice crystals if the temperature during the expansion drops to −35.0°C or lower and the air is super-saturated with respect to ice at this temperature. He observed that the ice crystals were formed even though the expansion ratio was far less than that necessary to produce the formation of liquid water drops. Cwilong observed that if "ordinary atmospheric air" is used, ice crystals form at −27°C, and that if the air is "artificially contaminated with tobacco smoke," the limiting temperature is −23°C. He found "crystalline dust" to be no more active than "amorphous dust" in producing ice crystals.

Following Schaefer's work, experiments were made using Schaefer's apparatus and technique to see whether particles of substances very similar to ice in crystal structure might not serve as nuclei for ice formation at temperatures closer to the freezing point.

A search was made through x-ray crystallographic data for substances resembling ice as closely as possible in crystal system, space group,

TABLE I.

| Substance | System | Space group | Lattice constant | |
|---|---|---|---|---|
| Ice | hex. | $D_{6h}^4$ | 4.535 | 7.41* |
| AgI | hex. ZnO | $C_{6v}^4$ | $4.585_6$ | 7.490** |
| $PbI_2$ | hex. | $D_{3d}^3$ | 4.54 | 8.86 |

* William H. Barnes, Proc. Roy. Soc. London **A125**, 670–693 (1929).
** N. H. Kolkmeijer, W. J. D. von Dobbenburgh, and H. A. Boekehoogen, Proc. Amsterdam 31, 1014–1027 (1928).

---

[1] V. J. Schaefer, "The production of ice crystals in a cloud of supercooled water droplets," Science **104**, 457–459 (1946).

[2] B. M. Cwilong, Nature **155**, 361–362 (1945).

and dimensions of unit cell. Two substances were chosen which are listed in Table I along with ice for comparison.

These crystalline substances were tested to see if they would act as nuclei for the formation of ice when dusted as a powder into a super-cooled water cloud in Schaefer's apparatus at a temperature of −20°C.

Silver iodide was without apparent effect; however, the introduction of lead iodide powder caused the formation of a number of ice crystals in the super-cooled cloud. It appeared that far fewer ice crystals formed than would have been expected from the number of particles of lead iodide introduced.

The supposition that the lead iodide particles acted as nuclei because one of the dimensions of its unit cell was almost the same as that for ice was discarded when V. J. Schaefer discovered that iodiform crystals and iodine vapor behave in a similar way. The orthorhombic structure of iodine bears little relation to the structure of ice.

Later, investigations were made on the nucleating effect of smokes produced by electric sparks between electrodes of various metals. It was discovered that a single spark between silver electrodes in the presence of iodine vapor produced many thousands of times as many ice nuclei as lead iodide particles or iodine vapor alone. It was then found that silver iodide smoke, produced by heating silver iodide on a hot filament or by dispersing it in a flame, produced enormous numbers of nuclei. The failure of the initial experiment using silver iodide powder is attributed to the fact that the sample used was badly contaminated with soluble salts, a fact not known at the time. Powdered silver iodide without this impurity has since been found to serve as nuclei for ice formation.

It is believed that silver iodide acts as a very effective nucleus for the formation of ice crystals because it very closely resembles ice in crystal structure. It can be seen that both dimensions of the unit cell of ice and silver iodide are the same to within about one percent.

According to Dr. D. Harker of this laboratory, the arrangements of the atoms in the unit cells of ice and of silver iodide are almost identical despite their different space groups. The structure of ice is the same as that of silver iodide with the oxygen atoms occupying positions corresponding to the silver and iodine atoms. In silver iodide, each silver atom is bonded tetrahedrally to four iodine atoms. In ice each oxygen atom is bonded tetrahedrally, through hydrogen bridges, to four oxygen atoms.

Iodine vapor alone, or fine silver particles alone are without appreciable effect in forming ice crystals at a temperature of −20°C. If the experiment is carried out using iodine vapor alone in a system closed to the atmosphere of the laboratory, iodine vapor soon loses its property of nucleating a super-cooled cloud. It is believed that iodine compounds and iodine vapor act to produce nuclei by reacting to form silver iodide with minute traces of silver in the laboratory atmosphere. Such traces of silver might be caused by sparks from electrical equipment using contacts made of silver or copper that contains silver. Exceedingly small amounts of silver introduced into a super-cooled cloud in the presence of iodine vapor cause the formation of very large numbers of nuclei. One breath of air blown over a silver wire heated red will produce many millions of ice nuclei if it is introduced into a super-cooled cloud containing a small amount of iodine vapor.

Experiments are being made on methods for the production of silver iodide smokes consisting of very large numbers of very small particles. The most effective method found thus far for the production of such smokes is the following. A cotton string coated with silver iodide is fed at a fixed rate into an oxyhydrogen flame which vaporizes the silver iodide. A few inches away from the point at which the silver iodide is introduced, the flame is rapidly quenched by blowing a strong jet of compressed air through it. The blast of air quickly cools and dilutes the vaporized silver iodide in the flame so rapidly that it condenses to form an invisible smoke of very small silver iodide particles. Measurements have been made which show that a generator of this sort produces ice nuclei at the rate of approximately $10^{13}$ per second with the consumption of one milligram of silver iodide per second. Mr. E. F. Fullam has examined particles of this smoke with the electron microscope and found that they are of the order of 100 Angstrom units in diameter. One milligram, it can be computed, should yield about $3\times 10^{14}$ particles. The discrepancy between the two values may be caused by the fact that a portion of the silver iodide introduced into the flame may form a few very large particles or may form particles which for some reason are not effective nuclei.

It has been found that the effectiveness of the silver iodide particles as ice nuclei depends to a certain extent on the size of the particles. The smoke made by the oxyhydrogen flame technique has a particle size of the order of 100A diameter. This smoke does not cause large numbers of crystals to form until the temperature is −8°C or lower. Smokes made by vaporizing silver iodide from a hot wire have a particle size of the order of one micron and are effective at −4°C or below. A piece of silver iodide several millimeters in diameter placed in water in a test tube makes it difficult to super-cool the water to temperatures below about −3.5°C.

Recently the effect of silver iodide smoke was tried near Schenectady on a thin layer of supercooled fog at a temperature of −4°C. The smoke was produced by evaporating silver iodide on the surface of an electrically heated wire coil. The liquid water fog was transformed into small ice crystals for a distance of at least 150 feet from the point where the smoke was generated.

---

# Waves in Elastic Tubes: Velocity of the Pulse Wave in Large Arteries

ALLEN L. KING
*Dartmouth College, Hanover, New Hampshire*

(Received February 28, 1947)

An equation for waves in elastic tubes is developed and applied to cylindrical tubes with Hookian and with elastomeric walls. The former application yields the Moens-Korteweg formula, which has been found inadequate. The latter application leads to a rather complicated equation for the pulse-wave velocity.

Values of pulse-wave velocities are computed for the thoracic aorta by means of the foregoing equation, and the results are compared with measured mean velocities for the entire aorta. Graphs and tables are given, so that a graphical analysis of this velocity equation can be applied to any large artery; and the method is illustrated by computing the pulse-wave velocity in the left common carotid. The relation $l\div\int_0^l ds/v$, for the mean velocity over a tube of length $l$, is shown to be valid for the aorta and large arteries.

IN a series of papers Hamilton, Remington, and Dow have published an extensive and critical survey of previous work and of their own important contributions to the study of propagation velocities for the arterial pulse wave[1] and its relation to stroke volume[2] and the cardiac ejection curve.[3] They found their results to be in accord with those of previous investigators. At best, only qualitative agreement was found between measured velocities and those computed from the Moens-Korteweg formula[4]

$$v=F(Ee_0/2\rho r')^{\frac{1}{2}}. \qquad (1)$$

In this relation $E$ is Young's modulus of the wall material, $e_0$ is the thickness of the undistended wall, $r'$ is the internal equilibrium radius of the tube, and $\rho$ is the density of the fluid within the tube. Moens' factor $F$ is an empirical dimension-

---

[1] W. F. Hamilton, J. W. Remington, and P. Dow, Am. J. Physiol. **144**, 521 (1945).

[2] J. W. Remington, W. F. Hamilton, and P. Dow, Am. J. Physiol. **144**, 536 (1945).

[3] J. W. Remington, and W. F. Hamilton, Am. J. Physiol. **144**, 546 (1945).

[4] Moens, A. I. *Die pulsekurve* (E. J. Brill, Leiden, 1878) p. 90; D. S. Korteweg Ann. d. Physik u. Chem. **241**, 525 (1878). Dr. H. J. Ralston of the College of Physicians and Surgeons of San Francisco has called my attention to a paper by Thomas Young [Trans. Roy. Soc. London **98**, 164 (1808)], wherein under the title, "Of the propagation of an impulse through an elastic tube," a formula for pulse-wave velocity is given, which can be transformed into that commonly ascribed to Korteweg.

less constant to which Moens assigned a value of 0.9, as a result of experiments with standing waves in elastic tubes. Hamilton and co-workers found the same value of Moens' factor for Gooch tubing; but for aortas they obtained better agreement with a value of 0.6–0.7 for the dog and 0.8 for the human; at diastolic pressure levels above 70 mm Hg. At lower pressure levels Moens' factor is not constant, and there is not even qualitative agreement.

In this paper a general wave equation is derived for the case of a homogeneous, non-viscous, and incompressible fluid in a thin-walled elastic tube. The Moens-Korteweg formula, with $F$ equal to unity, is readily obtained by introducing a simple version of the pressure-radius relation for a Hookian elastic tube with walls of negligible thickness. Aortas, however, have elastic properties similar to those of rubber tubes and, therefore, the pressure-radius relation for tubes with elastomeric walls should be used.[5] Pulse-wave velocities computed from the new equation are in good agreement with measured values. Even so, it should be realized that arterial pulsations are larger and vessel walls are thicker than assumed here. And, furthermore, blood is neither homogeneous nor non-viscous.

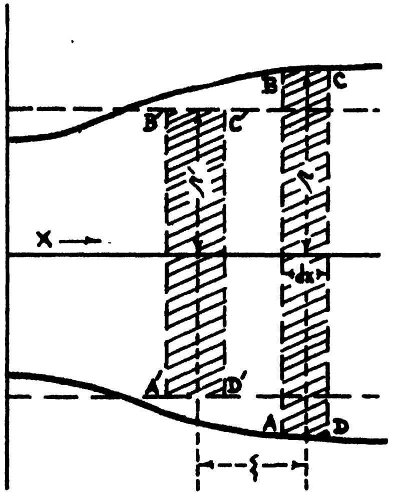


FIG. 1. Diagram of a portion of an elastic tube over which a wave is traveling.

## THE WAVE EQUATION

A homogeneous non-viscous incompressible fluid flows through a cylindrical tube with an elastic wall. The wall is sufficiently thin so that its inertia can be neglected in comparison with that of the fluid. The speed of a transverse wave down the tube is desired.

Let a thin element of fluid $ABCD$ of length $dx$ and radius $r$ be displaced a distance $\xi$ from its undisturbed position $A'B'C'D'$ when a wave passes along the tube (Fig. 1). With no wave in the tube the radius is $r'$, and the internal pressure is $p'$. At all times the external pressure is $p_0$.

The pressure on the fluid arises from the distension of the tube wall and is a function of the radius. Let the instantaneous excess pressure $P$ at $AB$ equal $(p-p')$, where $p$ is the instantaneous absolute pressure at $AB$. Then the excess force on surface $AB$ is $\pi r^2 P$, and the excess force on surface $CD$ is $\pi r^2 P+(\partial/\partial x)(\pi r^2 P)\cdot dx$. The force accelerating the element $ABCD$ in the positive $x$-direction, therefore, is $-(\partial/\partial x)(\pi r^2 P)dx$. But, the mass of this fluid element is $\pi r^2 \rho dx$; so that the equation of motion can be written

$$-\frac{\partial}{\partial x}(r^2 P)=\rho r^2\frac{\partial^2\xi}{\partial t^2}. \qquad (2)$$

The thickness of the element at $A'B'C'D'$ is $(1-\partial\xi/\partial x)dx$ and, since the fluid is incompressible,

$$r^2=r'^2(1-\partial\xi/\partial x). \qquad (3)$$

Equation (2) can now be written

$$\frac{\partial^2\xi}{\partial t^2}=\frac{1}{\rho}\left\{P(1+\partial\xi/\partial x)\frac{\partial^2\xi}{\partial x^2}-\frac{\partial P}{\partial x}\right\}. \qquad (4)$$

This is the equation for the pulse-wave.

## TUBES WITH HOOKIAN WALLS

For tubes with walls having a constant value of elastic modulus and following Hooke's law, Korteweg used the simple relation in which the wall thickness is assumed not to change. For this case the instantaneous excess pressure

$$\begin{aligned}P&=Ee_0(1/r'-1/r)\\&=-(Ee_0/2r')(\partial\xi/\partial x).\end{aligned} \qquad (5)$$

On inserting this value of $P$ in the general equa-

---

[5] A. L. King, J. App. Phys. 17, 501 (1946).

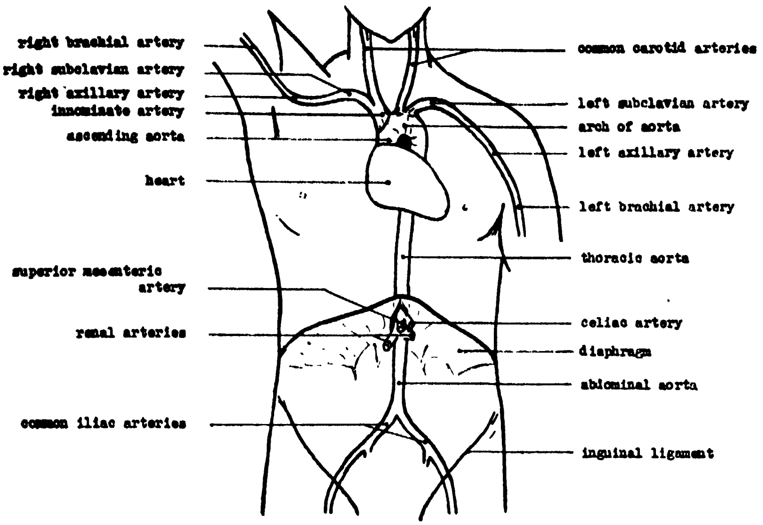


FIG. 2. Schematic drawing showing the approximate positions of the aorta and some of its branches.

tion for the pulse-wave and on neglecting terms involving the square of $\partial\xi/\partial x$, the wave equation,

$$\frac{\partial^2\xi}{\partial t^2}=(Ee_0/2\rho r')\frac{\partial^2\xi}{\partial x^2}, \qquad (6)$$

is obtained. The velocity of the wave is

$$v=(Ee_0/2\rho r')^{\frac{1}{2}}, \qquad (7)$$

namely, the Moens-Korteweg formula with $F$ equal to unity.

In general, the thickness of the tube wall does not remain constant. If this effect is taken into account, the instantaneous excess pressure is given by

$$\begin{aligned}P&=2e_0E(r-r')/(2-\sigma)r_0^2\\&=-\{Ee_0r'/(2-\sigma)r_0^2\}(\partial\xi/\partial x),\end{aligned} \qquad (8)$$

where $\sigma$ is Poisson's ratio of the wall material. When this value of $P$ is used in Eq. (4) and terms involving the square of $\partial\xi/\partial x$ are neglected, the wave velocity comes out to be

$$v=\{Ee_0r'/\rho r_0^2(2-\sigma)\}^{\frac{1}{2}}. \qquad (9)$$

This relation for pulse-wave velocity yields a somewhat higher value than that obtained from Eq. (7), since $r'$ is greater than $r_0$ and $\sigma$ ordinarily is positive.

TABLE I. Values of $A$ and $\beta$ for several age groups.

| Age range (years) | 20–24 | 29–31 | 36–42 | 47–52 | 71–78 |
|---|---|---|---|---|---|
| $A$ (mm Hg) | 88 | 74 | 70 | 65 | 44.5 |
| $\beta$ | 0.302 | 0.415 | 0.462 | 0.510 | 0.640 |

## TUBES WITH ELASTOMERIC WALLS

Elastomers are rubber-like substances containing long chains of molecules. They are highly resilient and have the unusual thermoelastic property of contracting on heating, when they are under tension. Living tissues have these characteristics and, in particular, the walls of blood vessels may be considered elastomeric.

For a thin-walled elastomeric tube the instantaneous excess pressure is

$$\begin{aligned}P=A[(r_0/r)^{\frac{1}{2}}\{\mathfrak{L}^{-1}(\beta r/r_0)/\mathfrak{L}^{-1}(\beta)\}\\-(r_0/r')^{\frac{1}{2}}\{\mathfrak{L}^{-1}(\beta r'/r_0)/\mathfrak{L}^{-1}(\beta)\}\\+(r_0/r')^{\frac{1}{2}}-(r_0/r)^{\frac{1}{2}}],\end{aligned} \qquad (10)$$

where $A$ represents $e_0p_0/2r_0$, and $\beta$ is equal to the ratio of the circumference of the tube to the maximum length of a molecular chain.[5] For arteries $\beta$ is an age-dependent parameter. The functions $\mathfrak{L}^{-1}(\beta)$, $\mathfrak{L}^{-1}(\beta r/r_0)$, and $\mathfrak{L}^{-1}(\beta r'/r_0)$ are inverse Langevin functions.

For small variations in the radius, as the wave progresses along the tube, $\mathfrak{L}^{-1}(\beta r/r_0)$ may be expanded about $r'$, and only the first two terms need be retained. The result is

$$\mathfrak{L}^{-1}(\beta r/r_0)=u'\{1-\tfrac{1}{2}(Br'/r_0)(\partial\xi/\partial x)\}, \qquad (11)$$

where

$$B=\beta u'\sinh^2 u'/(\sinh^2 u'-u'^2) \tag{12}$$

and

$$u'=\mathfrak{L}^{-1}(\beta r'/r_0). \tag{13}$$

Eq. (10) now reduces to

$$P=\tfrac{1}{4}A(r_0/r')^{\frac{1}{3}}\{(1-2Br'/r_0)(u'/u_0)\\ -3(r_0/r')\}(\partial\xi/\partial x), \tag{14}$$

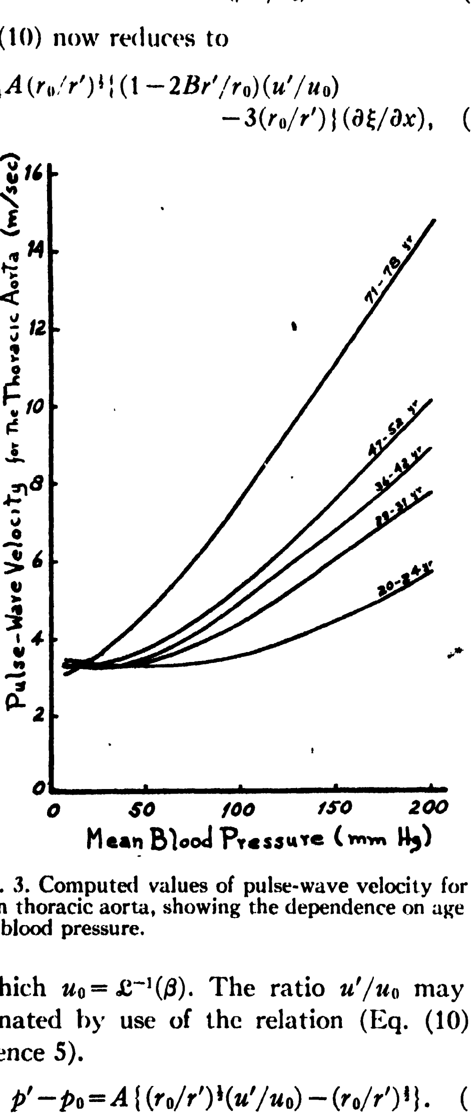


FIG. 3. Computed values of pulse-wave velocity for the human thoracic aorta, showing the dependence on age and mean blood pressure.

in which $u_0=\mathfrak{L}^{-1}(\beta)$. The ratio $u'/u_0$ may be eliminated by use of the relation (Eq. (10) in reference 5).

$$p'-p_0=A\{(r_0/r')^{\frac{1}{3}}(u'/u_0)-(r_0/r')^{\frac{4}{3}}\}. \tag{15}$$

Insert the resultant expression for $P$ in Eq. (4), neglect terms involving the square of $\partial\xi/\partial x$, and solve for the velocity of propagation of a pulse-wave in the elastomeric tube. The velocity is given by the equation

$$v^2=\tfrac{1}{4}(A/\rho)[(2Br'/r_0-1)\\ \times\{(p'-p_0)/A+(r_0/r')^{\frac{4}{3}}\}+3(r_0/r')^{\frac{4}{3}}]. \tag{16}$$

## PULSE-WAVE VELOCITIES IN THE HUMAN AORTA

Equation (16) for pulse-wave velocity may be applied to blood vessels. Note that the velocity increases with a rise in mean pressure within the tube. Also it increases with age, for the parameter $\beta$ in the expression for $B$ is age-dependent. Furthermore, it is higher when the ratio of wall thickness to internal diameter is larger, primarily through the factor $A$. All these conclusions are in qualitative agreement with observations on arteries of dog and man.

For a quantitative study, pulse-wave velocities are computed for that part of a human thoracic aorta just beyond the arch (Fig. 2). Values of $A$ and $\beta$ for several age groups were evaluated elsewhere[5] and are given in Table I. The specific gravity of blood is assumed to equal 1.057 for all ages.[6] The resulting computed velocities at pressures of 7.5, 50, 100, 150 and 200 mm Hg are gathered in Table II and are plotted in Fig. 3. Near zero pressure, the velocity is seen to decrease as the aorta ages. At the normal ranges of aortic pressures, however, there is a general increase of velocity with age. This increase becomes much pronounced in the hypertensive range. As noted in Fig. 3, the pulse-wave velocity is very nearly a linear function of pressure above 120 mm Hg.

Ordinarily, experimental determinations of pulse-wave velocities are made over long arterial paths. For instance, Hallock[7] made an extensive statistical study of aortic velocities in nearly 500 humans ranging in age from 5 to 85 years. His values of mean velocities are based on measurements of differences in distances and propagation times for pulse-waves traveling over the arch, thoracic and abdominal aortas, and the left iliac artery to the inguinal ligament and over the left common carotid artery (Fig. 2). To compute such mean velocities it would be necessary to know the velocity over every element of length along

TABLE II. Pulse-wave velocities for the thoracic aorta in meters per second.

| Age (years) | Pressure (mm Hg) 7.5 | 50 | 100 | 150 | 200 |
|---|---|---|---|---|---|
| 20–24 | 3.47 | 3.29 | 3.57 | 4.50 | 5.74 |
| 29–31 | 3.31 | 3.36 | 4.38 | 6.11 | 7.88 |
| 36–42 | 3.28 | 3.49 | 4.91 | 6.70 | 8.94 |
| 47–52 | 3.26 | 3.71 | 5.32 | 7.53 | 10.16 |
| 71–78 | 3.02 | 4.65 | 7.57 | 11.12 | 14.73 |

---

[6] W. F. Hamilton, Glasser's *Medical Physics* (Year Book Publishers, Chicago, 1944) p. 115.

[7] P. Hallock, Arch. Internal Medicine **54**, 770 (1934).

the path. In such cases, the mean velocity $\bar{v}$ is given by the relation

$$\frac{1}{\bar{v}}=\frac{1}{l}\int_0^l \frac{ds}{v}, \qquad (17)$$

where $l$ is the path length. If aging is uniform over the entire aorta, the ratio of mean velocity to the velocity in the thoracic aorta alone should be independent of age. Furthermore, the ratio should be greater than unity, since the pulse-wave velocity increases as the vessels become smaller in diameter and their walls thicken. On comparing the theoretical values of velocities for the thoracic aorta at a mean blood pressure of 100 mm Hg with Hallock's measured mean velocities, a constant empirical factor of 1.37 can be found. The approximate agreement between the adjusted theoretical values and Hallock's data is shown in Fig. 4.

### PULSE-WAVE VELOCITIES IN OTHER LARGE ARTERIES

Bazett, Cotton, Laplace, and Scott[8] published measured values of average pulse-wave velocities for paths from the heart, through the ascending aorta, arch, and innominate artery, to the subclavian artery, and also from the subclavian, through the axillary, to the branchial artery (Fig. 2). For the three men in the age-range of 20–24 years, whom they investigated, the mean values of these velocities were 3.6 and 5.4 meters/sec., respectively, with less than 10 percent deviation. To obtain the mean velocity between heart and brachial artery from these data, Eq. (17) may be modified. Since the distances from heart to subclavian artery and from subclavian to brachial artery, as described by Bazett and co-workers, are approximately equal, the reciprocal of the mean velocity between heart and brachial artery will be equal to half the sum of the reciprocals of the two measured velocities. This computation yields the value 4.3 meters/sec.

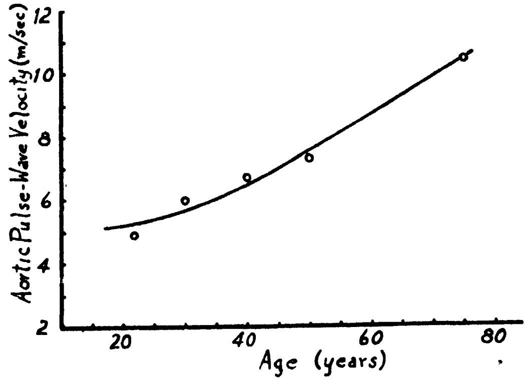


FIG. 4. The curve shows Hallock's values of mean pulse-wave velocities for the entire aorta as a function of age. The points were obtained by multiplying the theoretical values of velocity for the thoracic aorta alone, at a mean blood pressure of 100 mm Hg, by the empiric factor 1.37.

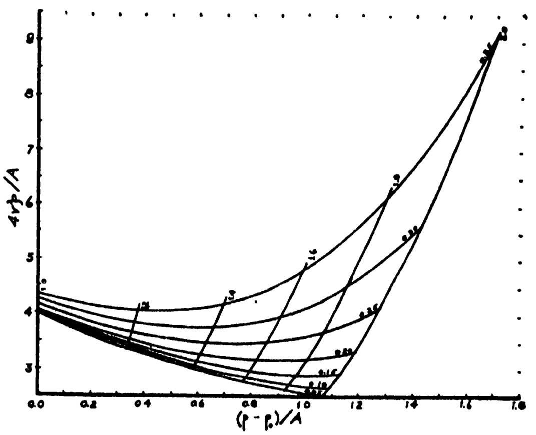


FIG. 5. Graph of $4v^2\rho/A$ as a function of $(p-p_0)/A$ for several values of $\beta$ and $r/r_0$. Values of $\beta$ range from 0.05 to 0.35 in steps of 0.05, and values of $r/r_0$ range from 1.0 to 2.0 in steps of 0.2.

The foregoing result may be compared with the observations of Bazett and Dreyer.[9] These investigators measured the pulse-wave velocity between heart and brachial artery of eight medical students with ages in the range 20–24 years. The mean velocity was equal to 4.1 meters/sec. at a mean blood pressure of 88 mm Hg with an average deviation of ±0.4 meter/sec.

Since the average dimensions of the arteries from aorta to brachial artery are equivalent to those of the common carotid, it may well be expected that the pulse-wave in the carotid should have a velocity of approximately 4 meters/sec. An extension of the analysis in the early part of this paper can be used to compute the velocity in any large artery. It will be applied to the common carotid artery.

---

[8] H. C. Bazett, F. S. Cotton, L. B. Laplace, and J. C. Scott, Am. J. Physiol. 113, 312 (1935).

[9] H. C. Bazett and N. B. Dreyer, Am. J. Physiol. 63, 94 (1922).

## GRAPHICAL ANALYSIS OF THE VELOCITY EQUATION

In Eqs. (15) and (16) $r'/r_0$ and $\beta$ may be considered parameters for the quantities $(p'-p_0)/A$ and $4\rho v^2/A$. There is shown in Fig. 5 a family of curves for which $\beta$ varies between 0.05 and 0.35 in steps of 0.05, and $r'/r_0$ varies between 1.0 and 2.0 in steps of 0.2. Other curves may be drawn for values of $\beta$ and $r'/r_0$ outside these ranges, but for the following computations the curves of Fig. 5 suffice.

In order to use these curves, first $A$ is evaluated from known values of external pressure and ratio of effective wall thickness to internal diameter of the undistended artery. Then from the mean blood pressure within the artery $(p'-p_0)/A$ is computed. A value for $\beta$ next is found by making the reasonable assumption that, for the walls of the large arteries within a given body, the molecular chains all have the same length. For all such arterial walls, therefore, $\beta$ is directly proportional to the internal diameter of the undistended vessel. Since values of $\beta$ already are known for the upper part of the thoracic aorta (Table I), $\beta$ for any other large artery may be found by the relation

$$\beta\,(\text{artery}) = \beta\,(\text{thoracic aorta}) \cdot \frac{d_0(\text{artery})}{d_0(\text{thoracic aorta})}. \qquad (18)$$

From these values of $\beta$ and $(p'-p_0)/A$ a value of $4\rho v^2/A$ may be read off the curves of Fig. 5 by interpolation and the pulse-wave velocity $v$ can then be found.

To illustrate the procedure, a value for the pulse-wave velocity in the common carotid artery of a 20–24 year old person will be computed. From Kani's data[10] the mean ratio of wall thickness to internal diameter is estimated to equal 0.22 for twenty-year olds. For an external pressure $p_0$ of 770 mm Hg, $A$ nearly equals 170 mm Hg. The average pressure within an artery near heart level is assumed to be the mean of the systolic and diastolic pressures. A fifteen percent error in this mean introduces less than two percent error in the computed velocity of the pulse-wave for the carotid artery. As noted earlier, the mean blood pressure for a 20-year old person nearly equals 90 mm Hg. The function $(p'-p_0)/A$, therefore, is equal to 90/170 or 0.53.

The ratio $d_0$(carotid)/$d_0$(thoracic aorta) for 20-year olds approximately equals one-third, so that for the carotid $\beta$ equals 0.10. From Fig. 5, the quantity $4\rho v^2/A$ is found to have the value 3.15 at values of $\beta$ and $(p'-p_0)/A$ equal to 0.10 and 0.53, respectively. But for whole blood the density $\rho$ equals 1.057 g/cc, and here $A$ equals 170 mm Hg, or 226,000 dynes/cm². From these data $v$ is computed to be 4.1 meters/sec.

The foregoing method may be modified so as to obtain values of $\beta$ from measurements of pulse-wave velocity and mean blood pressure.

No doubt other factors than those considered in this development may influence pulse-wave velocities in living systems. The experiments reported by Hamilton and co-workers[1] in their study of this problem, however, suggest that the propagation of a pulse wave along the aorta and large arteries depends almost completely on the mechanical properties of the intact living artery.

---

[10] I. Kani, Virchow's Arch. **201**, 45 (1910); see also A. L. King, Science **105**, 127 (1947) for comments on Kani's data.

# An Electronic Computer for X-Ray Crystal Structure Analyses*

R. PEPINSKY
*Department of Physics, Alabama Polytechnic Institute, Auburn, Alabama*

(Received January 29, 1947)

An electronic synthesizer is described for determination of atomic positions in crystals. The synthesizer sums the two-dimensional Fourier series representing planar, centro-symmetric projections of electron densities in a crystal unit cell; and the projection is presented by a television scan on the screen of a cathode ray oscilloscope. The specific advantage of the device is the immediate observability of effects on the projection of alterations in signs of one or any number of Fourier coefficients.

A METHOD has been developed for rapid visual presentation of the positions of atoms in a crystal structure by supplying x-ray diffraction data to an electronic computer coupled to a cathode ray oscillograph. The computer sums the two-dimensional Fourier series representing the projection on a lattice plane of the electron densities in a crystal unit cell; and the projection is presented by means of a television type of scan on the screen of the cathode ray tube. The method is based upon Huggin's modification[1] of Bragg's earlier photographic summation of simulated interference fringes[2]; in this respect it is related to the sand-deposition method of fringe summation reported by McLachlan and Champayne[3] in the December, 1946 issue of this journal. In contrast to these modifications of the Bragg method, however, accurate projections with various axial ratios and angles are readily permitted in the present computer.

Artificial interference fringes for an (*hk*0) projection are produced on the oscilloscope screen by applying to the CR tube grid intensity-modulating sinusoidal voltage signals of amplitude proportional to $F_{hk0}$ and frequency $f_{hk0}$, the latter being given by the relation:

$$f_0 = f_{hk0} = h\cdot f_V + k\cdot f_H, \tag{1}$$

where $f_V$ and $f_H$ are the vertical and horizontal sweep frequencies respectively. By way of illustration: $f_V$ may be a slow sweep, at 1 c.p.s., and $f_H$ a fast sweep at 1000 c.p.s., and the duty cycle of both sweeps practically 100 percent. One 1000-line scanning frame then appears on the screen each second. The signals for various (*hk*0) fringes are mixed before application to the grid, so that a two-dimensional pattern occurs for each frame. The fluorescent screen is a long-persistence type (3- to 5-second decay period) to permit visualization, and the pattern may be photographed for permanent recording.

Only patterns with centers of symmetry are attempted. The *signs* of the structure factors $F_{hk0}$, which in general are not directly determinable from the x-ray data, are introduced by having each signal available in 0 or 180° phase, and selecting one or the other of these by means of a double-throw switch for each term. Amplitudes are adjusted by means of individual potentiometers for each factor.

It will be seen from Eq. (1) that when $h=0$, so that $f_0$ is some integral multiple of the fast sweep, vertical fringes result on the screen. *Increasing* $f_0$ slightly, with the sweep frequencies unchanged, causes the fringes to tilt (but remain parallel to one another) in one direction, because the intensity-modulating signal *advances* in phase with respect to the fast sweep. *Decreasing* the modulating frequency produces parallel fringes oppositely tilted to the first case. Certain refinements of this simple procedure, involving the use of lower duty-cycle sweeps but of the same sweep speed as indicated above, are required to maintain accurate placement of the fringes. These and other details of the technique will be described and published elsewhere.

A fast sweep of 1000 c.p.s. in the horizontal direction and a slow of 1 c.p.s. vertically have been utilized, and precise synchronization of

---

* Presented in part before the American Society for X-Ray and Electron Diffraction, Lake George, New York, June 11, 1946.

[1] M. L. Huggins, Jr., J. Am. Chem. Soc. 63, 66 (1941).

[2] W. L. Bragg, Zeits. f. Krist. 70, 475 (1929).

[3] D. McLachlan, Jr., and E. P. Champaygne, Jr., J. App. Phys. 17, 1006 (1946).

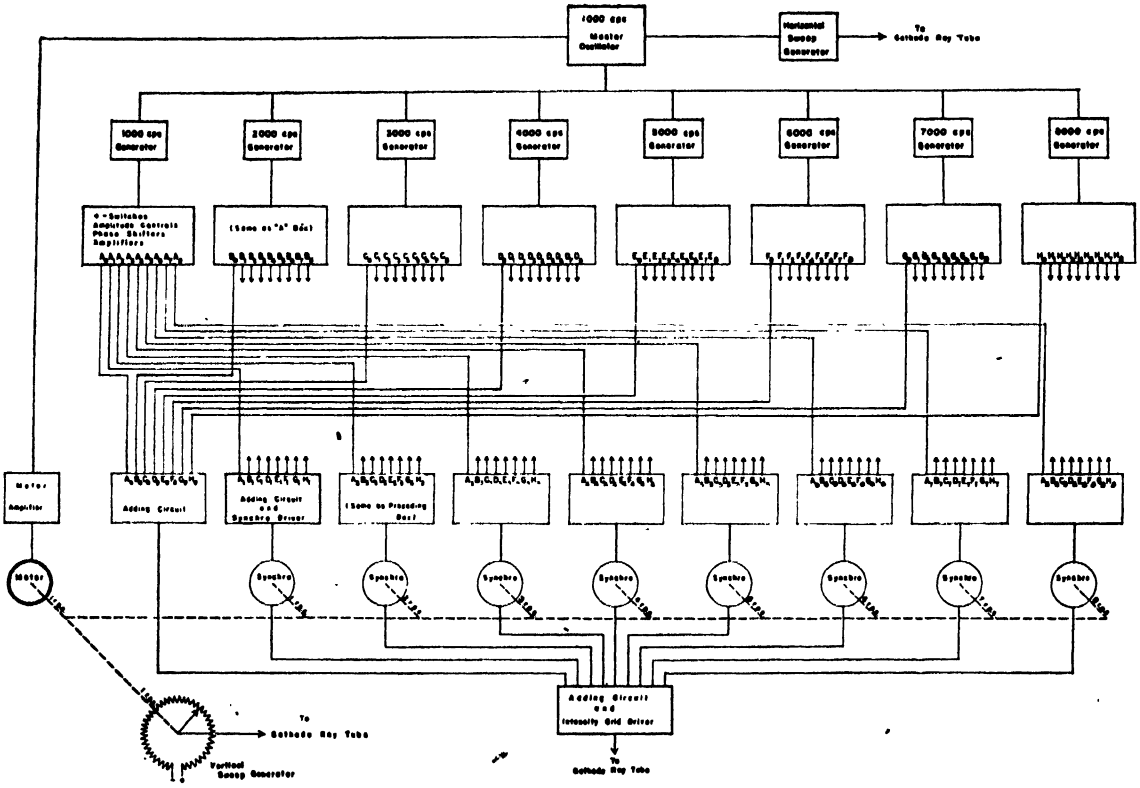


Fig. 1. Block diagram of Fourier synthesizer for (h, k) terms to (8, 8).

these with the grid signals, according to equation (1), is essential. All frequencies are derived from a 1000 c.p.s. base oscillator, which should be fairly frequency stable but need not be strictly so. The fast sweep is synchronized to this; the slow is obtainable through the use of a 1000 cycle motor with a 1 r.p.s. shaft, the motor being driven from a power amplifier.** A series of generators of frequencies 1000·k c.p.s. are synchronized by a heterodyning system to the base oscillator. Phase and amplitude-adjustment controls are available for each of these.

The key to the entire computer is the method for production of the frequencies 1000·k±h c.p.s. These must be accurately related to the base oscillator and to the fast and slow sweep speeds; no drifting whatever from the precise relation can be tolerated. The exactly-related frequencies are obtained as follows. Each 1000·k c.p.s. signal is fed to an integrating feed-back amplifier which produces a 90° phase-shifted signal of the original frequency. Application of these two signals—of equal amplitude and frequency but 90° phase-separation—to the stator windings of a two-phase selsyn produces a rotating electric field in the selsyn of rotational frequency 1000·k c.p.s. If the selsyn rotor remains stationary, a signal of frequency 1000·k c.p.s. appears on its leads. But if the rotor is rotated at h r.p.s. in a direction counter to the field, a signal of frequency 1000·k+h c.p.s. appears on the rotor leads. Rotation of the rotor at h r.p.s. in the *same* direction as the field rotation results in a signal of frequency 1000·k −h c.p.s. The rotor speeds are obtained by gearing in ratios 1:1, 1:2, ···1:h to the 1 c.p.s. shaft driven by the 1000-cycle motor.

All signals with the same h-value can be *mixed* and fed to a *single* selsyn, the rotor speed of which is h r.p.s. The number of selsyns required for the device is thus considerably less

---

** It will be demonstrated in a later paragraph that some of the synchronization here is redundant. Specifically, the motor need not be synchronized to the fast sweep signals and the 1000·k oscillators.

than the number of $F_{hk0}$ terms. The 90° phasing amplifiers are made rather frequency-insensitive, so that only one such amplifier is required for each selsyn; this again reduces the number of components. Because the output signal from the phasing amplifiers will vary with *large* variations in signal frequency, normalizing potentiometers are supplied for each of the 1000·k c.p.s. oscillators.

A 1000·k+h c.p.s. signal is produced, as described, by rotation of the selsyn rotor in a direction counter to the rotating field direction. It is possible to obtain both a 1000·k+h c.p.s. and a 1000·k−h c.p.s. signal from a *single* selsyn by producing rotating fields in *opposite* directions in the same selsyn stator. To accomplish this it is necessary only to alter the phase of one 90° component of one or the other of the 1000·k c.p.s. signals by 180°. (In general, the amplitude of the 1000·k c.p.s. signal which will become a 1000·k+h c.p.s. frequency is different from the 1000·k c.p.s. signal which will become the 1000·k−h c.p.s. frequency, the amplitudes being controlled by the $(h, k, 0)$ and $(h, \bar{k}, 0)$ potentiometers, respectively.) Thus, for a matrix of (20, ±20) terms, only twenty selsyns are required.

A partial block diagram for a system of $F_{hk0}$ terms up to (8, 8, 0) is shown in Fig. 1. The use of lower than 100 percent duty-cycle sweeps has not been indicated in this diagram, but such modifications do not alter the general operation of the system. The illustrated system achieves completely locked-in synchronization of all signals. Once the 1000·k c.p.s. signals have been phased so that their maxima occur at the start of the fast sweep, no drifting of fringes is possible.

A sample fringe, corresponding to $F_{\bar{1}70}$, is illustrated in Fig. 2. The fast sweep, at 1000 c.p.s., is from left to right in the figure; and the slow sweep, at 1 c.p.s., is from top to bottom. The grid signal here has the frequency 6999 c.p.s., as required by the formula:

$$f_{\bar{1}70} = 1\cdot h + 1000\cdot k = 1\cdot -1 + 1000\cdot 7 = 6999 \text{ c.p.s.}$$

The resulting pattern ordinarily will not show any discrete horizontal lines, because the distance between adjacent fast sweep traces is less than the breadth of the electron beam at its finest focus. In order to indicate the direction of the fast sweep here, an additional 60 c.p.s. modulation has been applied to the CR tube grid.

It can be shown, as suggested in the footnote on page 3, that there is a redundancy in the synchronization employed here. It is true that all the 1000·k oscillators and the fast sweep must be synchronized—and this is readily achieved. It is also necessary that the slow sweep and the selsyn rotations be synchronized, and this is achieved through the use of the potentiometer driven from the (approximately) 1 c.p.s. shaft and with the selsyns driven from the shafts geared to this. But it is *not* necessary that the shaft speeds, at multiples of about 1 r.p.s., be synchronized with the 1000 c.p.s. oscillator. If the shaft speed slows down, more horizontal sweeps occur per frame, and the phase-shift of each fringe signal is reduced; but these two effects exactly compensate one another, and the tilt of each fringe remains unaltered. This simplifies the circuitry, and permits the use of a standard (60-cycle) motor.

The fast and slow sweep directions are ordinarily at right angles. By adding to the fast sweep signal a *component* of the slow sweep signal, the angle between the effective slow sweep direction and the fast sweep can be varied over a wide range. The slow sweep component is obtained from another linear potentiometer on the same 1 c.p.s. shaft as the main slow sweep potentiometer. The amplitude of the slow com-

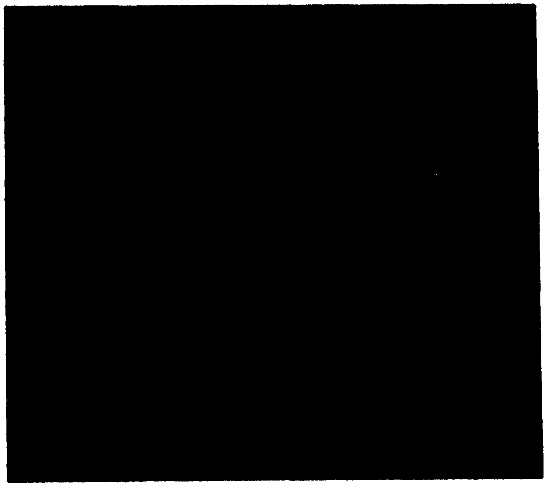


FIG. 2. $F_{\bar{1}70}$ fringe, with added 60 c.p.s. modulation to show fast sweep direction.

ponent is varied by means of the voltage applied across its potentiometer. Variation of the voltage across the main slow sweep potentiometer alters the axial ratio of the projection. Thus cell projections of any desired axial angle and ratio are attainable.

The computer can be built up in stages from one accommodating only a few $F_{hk0}$ terms to one including as large a range of terms as appears necessary. It is intended that terms to (20, ±20, 0) be incorporated into the final synthesizer.

The basic difficulty in general crystal structural analyses arises from the absence of information concerning the phases of the structure factors. Whenever a center of symmetry occurs in a structure or in some projection of the structure (e.g., a projection perpendicular to a two-fold symmetry axis is centro-symmetric), the factors involved in an electron density map for the projection are known except for their algebraic sign. Some partial information as to these signs may be available from auxiliary information (atomic radii and bond angles as determined from related structures, Patterson projections showing interatomic vectors, optical data indicating group or molecular orientations, etc.); and in special cases the signs may be directly determinable.[4] In general, however, trial and error procedures are required. Since the order of from one hundred to several hundred factors are involved in a projection, the number of permutations of signs is overwhelming; and when tedious computations must be carried through each time a change is made, without a great deal of guidance in assignment of the signs structure determinations become unfeasible.

In order to reduce the work involved in the calculations, a good number of computational aids have been developed for structural analyses. These will not be reviewed here. It suffices to say that none of them meet in a satisfactory manner the peculiar requirements of this science, that alterations of signs of the factors be enterable in a simple way into the computer, and the effects of the alterations be very rapidly observable.

It is precisely this requirement of structural analyses which the present synthesizer is designed to satisfy. The signs of individual factors can be introduced and altered at will by simply throwing a switch for each factor; and the "summation" of a two-dimensional series is accomplished within a few seconds after the data is fed into the machine. The pattern corresponding to a specific choice of signs remains on the oscilloscope screen as long as is desired. The effects of changes in signs of one or any number of factors appear almost instantly, and the new pattern remains on the screen until another change is introduced. (The amplitudes of individual factors are equally adjustable, with the effects of such changes just as rapidly observable.) *Thus for the first time systematic variations of structure factor signs, in trial and error procedures, become feasible in a general structure analysis.*

Another advantage of the oscillographic presentation is the ease with which the *contrast* and *background level* of the pattern is alterable. The former is controllable through amplitude control of the final grid-modulating signal; and the background level is controlled through variation of the CR tube grid bias. Since the grid-modulating signals are mixed before their application to the cathode-ray tube, a non-linear relation between the beam current and the light-output of the fluorescent screen does not lead to errors in positions of peaks in the electron-density map. The size of a projection is also controllable, and any section of a projection can be selected and enlarged at will.

The computer should not now be considered as a replacement for accurately calculated contour maps of electron densities in the final stages of a successful structural determination. It is intended to facilitate the deduction of general features of a structure and to avoid the tediousness of trial and error stages of the analysis. It is conceivable that with proper care a device of this sort will be capable of quantitative results, but such is not its prime purpose.

An earlier form of this instrument was described by the writer at the Lake George, New York, meeting of the American Society for X-Ray and Electron Diffraction, June 11, 1946, and oscilloscope fringe patterns were shown there. An extended discussion of the system will be published elsewhere.

---

[4] J. M. Robertson, Reports on Progress in Physics 4, 332 (1937).

# Resonant Cavities for Dielectric Measurements*

C. N. WORKS
*Research Laboratories, Westinghouse Electric Corporation, East Pittsburgh, Pennsylvania*

(Received March 3, 1947)

Fixed and variable length re-entrant resonant cavities designed for the measurement of dielectric constant and dissipation factor are described. These cavities operate in the frequency decade of $10^8$ to $10^9$ cycles per second, a region avoided by many experimenters because the frequency is too high for the application of circuit techniques and not high enough for the convenient use of coaxial lines or wave guides. The theoretical considerations in the design of these cavities are presented. The well-known susceptance variation method, widely used in the frequency range of $10^4$ to $10^8$ cycles per second, was extended to apply to these cavities; it yields a rapid measuring technique and very simple expressions for calculating the values of the dielectric properties. The performance of this apparatus is discussed and the results of the measurements of a few typical dielectrics are given.

## I. INTRODUCTION

DIELECTRIC constants and losses of materials had been measured even at ultra-high frequencies as early as 1895.[1] The early techniques were relatively crude compared with those developed within the last few years. Coaxial transmission line and wave guide methods, discussed by von Hippel,[2] Lamont,[3] Englund,[4] and Benoit,[5] in which a section of (usually resonent) coaxial line or wave guide is filled with dielectric, are used at present, especially at frequencies greater than 1000 megacycles per second.

The frequency band of 100–1000 megacycles has been avoided by most experimenters in dielectric measurements because the frequency is too high for the application of the usual circuit techniques and not high enough for the convenient use of the coaxial line or wave guide. The search for a simple and rapid method of making measurements in this frequency range has resulted in the development of two types of re-entrant cylindrical cavities in which small disk-shaped samples are used. This paper will be concerned with the description of these cavities as well as with their theory and performance.

## II. DESCRIPTION OF THE APPARATUS

### 1. The Fixed Cavity

This instrument, which is shown in Fig. 1, operates in the frequency range of 400–600 megacycles per second. It is similar to an early model developed for the frequency range of 100–300 megacycles,[6] except that as a result of

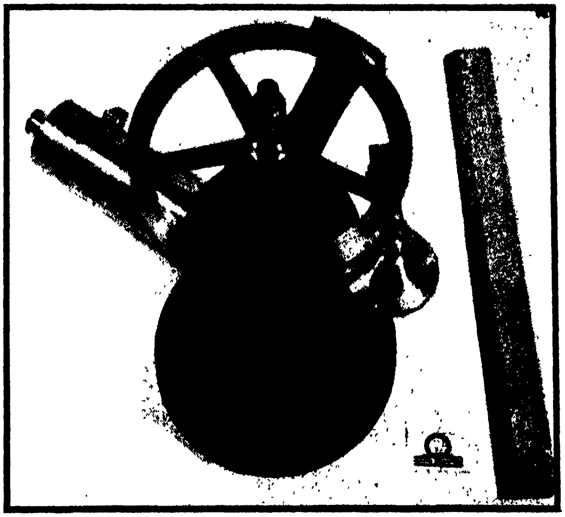


FIG. 1. The fixed cavity.

---

* The variable length cavity described in this paper was the subject of a thesis submitted to the Graduate School of the University of Pittsburgh, in partial fulfillment of the requirements of the M. S. degree.

[1] Drude, Wied Ann. **55**, 633 (1895); Zeits. f. physik. Chemie **23**, 267 (1897).

[2] S. Roberts and A. von Hippel, "New Method for Measuring Dielectric Constant and Loss in the Range of Centimeter Waves," J. App. Phys. **17**, 610–616 (1946).

[3] H. R. L. Lamont, "Theory of Resonance in Microwave Transmission Lines with Discontinuous Dielectric," Phil Mag. **29**, 521 (1940).

[4] C. R. Englund, "Dielectric Constants and Power Factors at Centimeter Wave Lengths," Bell Sys. Tech. J. **23**, 114–129 (1944).

[5] J. Benoit, J. de phys. et rad. **5**, 173–184, 203–216 (1944).

[6] C. N. Works, T. W. Dakin, and F. W. Boggs, "A Resonant Cavity Method for Measuring Dielectric Properties at Ultra High Frequencies," Proc. I.R.E. 245–254 (April 1945), or Trans. A.I.E.E. **63**, 1092–1098 (1944).

experience with this earlier instrument, a number of mechanical improvements have been made.

The sample required is a disk one inch in diameter and from 0.050 to 0.250 inch thick. It is inserted through the port near the bottom and placed between the re-entrant posts or electrodes for the measurement of its dielectric constant and dissipation factor.[7] This port is located in a portion of the cavity where the current is small. It is closed by a tapered plug. It has been found that the resonant frequency of the cavity is not altered by removal and replacement of the plug. The upper electrode is movable and is connected with the coaxial portion of the cavity by a sylphon bellows which eliminates all sliding contacts. The separation between these electrodes is controlled by a precision micrometer (Van Keuren Company) with a six-inch dial. The constructional details of this cavity are shown in Fig. 2.

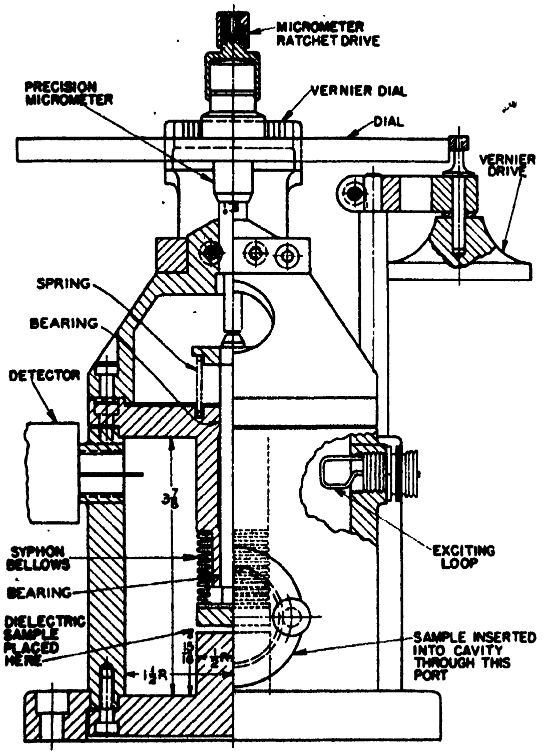
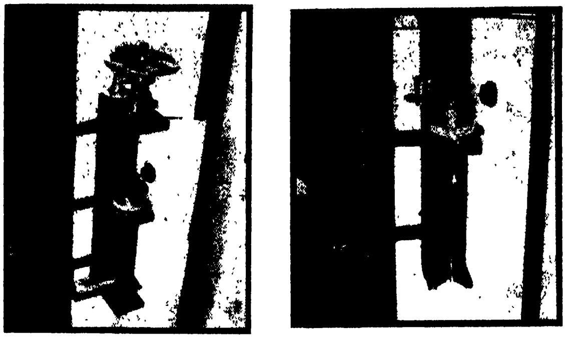


FIG. 2. 400–600-Mc dielectric measuring apparatus.



Assembled Cavity

Outer Cylinder of Cavity Removed Showing Internal Details

FIG. 3. 300–1000-Mc variable length re-entrant resonant cavity for measuring dielectric properties of materials.

Energy is coupled into the cavity by a loop as shown. By varying the size and position of this loop, the measured $Q$[8] of the cavity may be varied from less than 50 to about 2000. The detector consists of two type 1N21 crystals (Sylvania) side by side with ends of opposite polarity connected to the probe.

The fixed cavity has the limitation that the frequency at which the measurement of a given sample is made depends on the dielectric constant and dimensions of the sample. By a reasonable variation in the thickness of a sample, however, measurements may be made over the frequency band previously designated for this cavity.

## 2. The Variable Cavity

Because there may be cases where it is desirable to select the exact frequency at which a measurement is to be made, and also to obtain a single instrument useful over a wide band of frequencies, a variable length re-entrant cavity was designed for a frequency range of 300–1000 megacycles per second. Photographs of this instrument are shown in Fig. 3. Details of the cavity proper, which are not evident from these illustrations, appear in Fig. 4.

The length of the cavity is controlled by the position of a shorting plug. It might be emphasized that in this method variation in contact does not introduce an error in the measurements. The plug is driven by a threaded tube which is restrained from turning by a key on the extension

---

[7] Dissipation factor ($\tan\delta$) is the tangent of the loss angle. This angle is the complement of the phase angle whose cosine is the power factor. Therefore, for the small loss angles usually found in good dielectrics, the dissipation factor is equal to the power factor for all practical purposes.

[8] $Q$ is usually defined as the ratio of the amount of energy stored per cycle to the amount of energy dissipated per cycle. $Q$ is thus the reciprocal of the dissipation factor.

of the center post. This tube is propelled by a nut to which the dial (see Fig. 3) for indicating the length of the cavity is attached. A vernier drive is also provided.

The separation of the electrodes is controlled in the same manner as in the fixed cavity. In fact, all other details of the cavity are similar to the fixed cavity except that the former is smaller in diameter and has smaller electrodes, as indicated in Fig. 4. This smaller diameter results in a lower $Q$. By adjusting the coupling, the measured $Q$ may be caused to vary from less than 50 to about 1000. This maximum value of $Q$ still gives good sensitivity in measuring the dissipation factor of low loss materials. The sample required is a small disk 0.480″ or less in diameter, and from 0.040″ to 0.125″ thick. The auxiliary equipment used with these cavities is described in reference 6.

## III. THEORY

### 1. General Considerations in the Design of the Cavities

The well-known susceptance variation method[9,10] for the measurement of the dielectric constant and dissipation factor in the frequency range $10^4$ to $10^8$ cycles per second has a number of advantages. It combines a simple and rapid measuring technique with simple relations for rapid computation of results with an accuracy greater than that obtainable by most other methods. It has been found possible to apply this technique to measurements at higher frequencies. Expressions were derived for the dielectric constant and loss of samples, measured by placing them in the re-entrant end of a cavity and using a measuring procedure similar to that of the conventional susceptance variation technique. These are identical, except for an additional correction term, with those used in the usual application of this technique at radio frequencies, provided the dimensions of the sample are so restricted that it can be considered as a lumped capacitance.

If the sensitivity to low loss samples is to be satisfactory, the $Q$ of the cavity must be high. The expression for the theoretical $Q$ of such cavities is[11]

$$Q=4\pi(G/10)^{\frac{1}{2}}b\frac{\log_e b/a}{b/a+1}(f)^{\frac{1}{2}}\times 10^{-4}, \qquad (1)$$

where $G$ is the conductivity of metal covering inside of cavity in mhos per centimeter ($5.8\times 10^5$ for copper), $b$ is the inside radius of the outer conductor of the cavity in centimeters, $a$ is the outside radius of inner conductor of cavity in centimeters, and $f$ is the frequency in cycles per second. The optimum $Q$ is obtained with a ratio of $b/a$ of about 3.6. Examination of the curve[12] of $Q$ *versus* the ratio $b/a$ shows that this curve is fairly flat for values of $b/a$ between 3 and 4, and falls off rapidly on either side of these values. In the frequency band being considered, a theo-

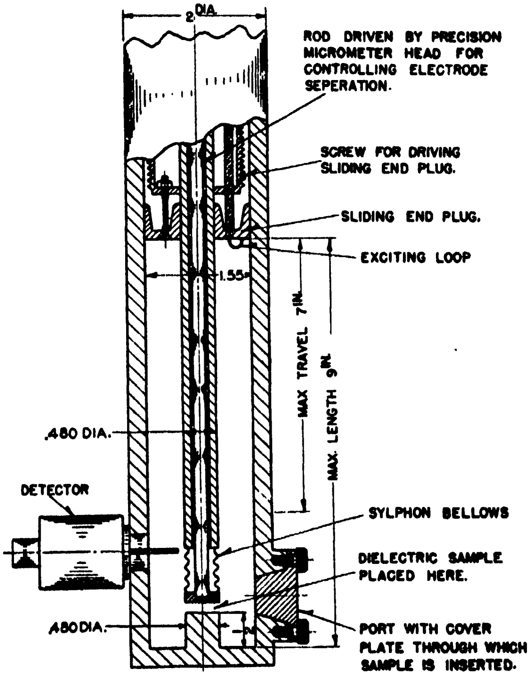

2 DIA.
ROD DRIVEN BY PRECISION MICROMETER HEAD FOR CONTROLLING ELECTRODE SEPERATION.
SCREW FOR DRIVING SLIDING END PLUG.
SLIDING END PLUG.
EXCITING LOOP
1.55
MAX TRAVEL 7 IN.
MAX. LENGTH 9 IN.
.480 DIA.
DETECTOR
SYLPHON BELLOWS
DIELECTRIC SAMPLE PLACED HERE.
.480 DIA.
PORT WITH COVER PLATE THROUGH WHICH SAMPLE IS INSERTED.

FIG. 4. Variable length re-entrant cavity.

---

[9] L. Hartshorn and W. H. Ward, "The measurement of the permittivity and power factor of dielectrics at frequencies from $10^4$ to $10^8$ cycles per second," J. Inst. Elec. Eng. **79**, 597–609 (1936).

[10] A.S.T.M. Committee D-9 "Tests for power factor and dielectric constant, susceptance variation method," *A.S.T.M. Standards on Electrical Insulating Materials* (The American Society for Testing Materials, Philadelphia, Pennsylvania, 1944), p. 341.

[11] Slater, *Microwave Transmission* (McGraw-Hill Book Company, Inc., New York).

[12] R. C. Miedke, "$Q$ for unloaded concentric transmission lines," Electronics 139–140 (September 1943).

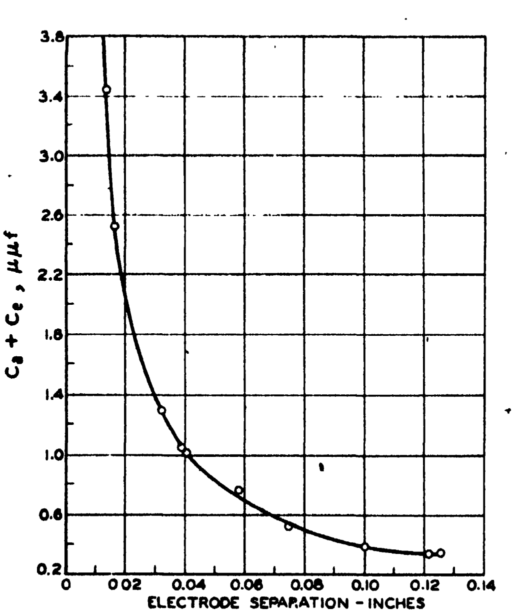


FIG. 5. Curve for the determination of $(C_{a2}+C_{e2})-(C_{a1}+C_{e1})$.

retical $Q$ of about 4000 may be obtained for values of $b$ as small as 2 cm. Because of energy absorbed by the probe and the effect of the coupling loop, the theoretical $Q$ is never obtained. Measured $Q$ of one-quarter to one-half of the theoretical $Q$ may be obtained under practical operating conditions.

Having established the restrictions imposed to obtain satisfactory $Q$ values, the next step is to determine the dimension of cavities, which resonate in the desired frequency band. The resonant frequency of the singly re-entrant cavity for the general case is treated by Hansen.[13] This case is so complex that it is at once apparent that it would be desirable to impose restrictions on the shape of the cavity which would result in simpler relations. E. U. Condon[14] gives an approximate treatment, which assumes that $b$ is small compared with a quarter wave-length. He also shows that this treatment corresponds to applying the standard engineering form of transmission line theory. Therefore, this restriction will be applied to these cavities, and transmission line formulae used to calculate the resonant frequency or resonant length. The resonant frequency of a short circuited loss-free transmission line with a capacitor at one end is given by the equation[11]

$$Z_0 \tan\frac{2\pi fS}{c}=\frac{1}{2\pi fC}, \tag{2}$$

where $Z_0$ is the characteristic impedance of coaxial part of cavity, $S$ is the length of internal conductor, $c$ is the velocity of light, and $C$ is capacitance between the ends of the re-entrant posts in the cavity at resonance. The resonant length of the cavities, as calculated from this relation, checked the resonant length as determined experimentally within two percent.

Experiments with a singly re-entrant cavity indicated that the fringing of the field is excessive, especially for very short gaps. Therefore, the doubly re-entrant cavities of Figs. 2 and 4 were selected in which one of the re-entrant posts is much shorter than the other. On first consideration, it would seem that the re-entrant posts should be the same length, giving a symmetrical cavity which might be easier to analyze. However, the requirement of variable gap length would disturb this symmetry. It was observed that for small gap spacings (less than $0.02S$) and for the short post less than $0.25S$, the resonant frequency of the doubly re-entrant cavity varied

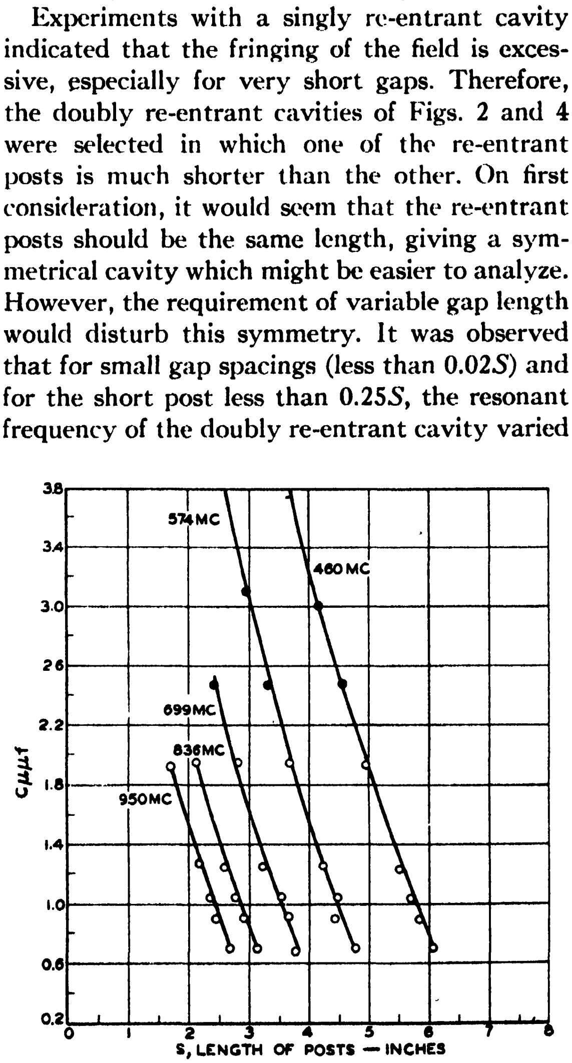


FIG. 6. Curves showing relation between geometric capacitance, $C$, between electrodes *versus* length of posts at constant resonant frequency.

---

[13] W. W. Hansen, "On the resonant frequency of closed concentric lines," J. App. Phys. 10, 38–45 (1939).

[14] E. U. Condon, "Principles of Microwave Radio," Rev. Mod. Phys. 14, 341–389 (1942).

very slightly from that of a singly re-entrant cavity of the same total length and same gap spacing. This observation confirms the experimental results obtained by Barrow and Mieher.[15]

The lowest mode of oscillation of this cavity corresponds to that of a conventional coaxial line short circuited at one end and capacitatively loaded at the other. It is usually designated $TM_{0,0,p}$ where $p$ is less than unity. Since the diameter of our cavities has been restricted so that the resonant frequency could be calculated by transmission line methods, it is unlikely that higher modes could exist. Calculation of the cut-off frequency for the next higher mode verified this.

## 2. Expressions for the Dielectric Constant

By definition, the dielectric constant, $e'$, of any dielectric sample is given by

$$e' = C_x/C_a, \tag{3}$$

where $C_x$ is the capacitance of the sample, and $C_a$ is its equivalent air capacitance. The capacitance of the sample, $C_x$, is given by the equation[6]

$$C_x = (C_{a2} + C_{e2}) - (C_{a1} + C_{e1}) + C_{a1} - \frac{dC}{dS}\Delta S. \tag{4}$$

Except for the last term, this is the same expression used in the conventional susceptance variation method.[10]

In applying this equation to the cavities, the more complicated case of the variable length cavity will be considered first. The capacitance of the electrode arrangement of two parallel cylinder faces is determined at lower frequencies by calibration against a standard capacitor in the case of the conventional susceptance variation circuit. In the frequency band being considered here, neither a standard capacitor nor the technique of comparing it to an unknown is available. Therefore, the calibration was determined experimentally, using two materials having considerably different values of dielectric constant, polystyrene and Insanol (mica-filled lead borate). The dielectric properties of samples of these materials were known from measurements made at higher and lower frequencies by other independent methods.

From the measurement of these materials of known properties, the calibration curve of Fig. 5 was obtained. This corresponds to the correction for fringing usually applied in the case of the conventional susceptance variation circuit. By obtaining the relation of the capacitance between electrodes to the length of the posts at constant resonant frequency, the curves of Fig. 6 were obtained. The curve of Fig. 7 was then obtained by plotting the slope of these curves *versus* electrode separation. Multiplying the ordinate of this curve by $\Delta S$, the change in the gap spacing to retune the cavity to resonance upon removal of the sample, gives the last term of Eq. (4). This term is the correction for the small change in length of the coaxial part of the cavity which necessarily takes place when one of the posts is extended inward to diminish the gap.[16]

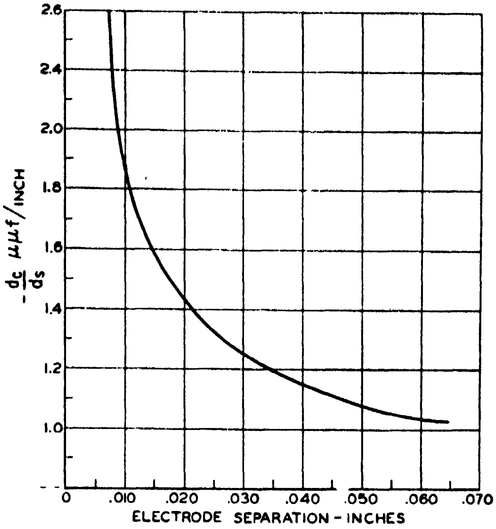


FIG. 7. Curve for the determination of $dC/dS$ for any electrode separation.

Now Eqs. (3) and (4) may be solved as follows to obtain the value of the dielectric constant of any material. The quantities $(C_{a2}+C_{e2})-(C_{a1}+C_{e1})$, and $-dC/dS$ are read from the curves in Figs. 5 and 7, respectively. The value of $C_{a2}+C_{e2}$

---

[15] W. L. Barrow and W. W. Mieher, "Natural Oscillations of Electrical Cavity Resonators," Proc. I.R.E. **28**, 184–191 (April 1940).

[16] The reader who may wish to perform the experiments described should refer to the thesis "A Variable Length Re-entrant Resonant Cavity for the Continuous Measurement of Dielectric Properties from 300–1000 Mc," University of Pittsburgh, 1946, for a detailed discussion of the method used to obtain Figs. 5, 6, 7.

is read from the ordinate of Fig. 5, corresponding to the value of electrode separation with the sample removed. Similarly, the value of $C_{a1}+C_{e1}$ is found from the separation with the sample in place.

In the case of the fixed cavity, samples of various thicknesses, whose dielectric constants were accurately known, were measured over the frequency range of 400–600 megacycles per second. Using the computed theoretical values for the correction term, $(dC/dS)\Delta S$, the measurements all checked the known values within one percent. This indicated that fringing could be neglected in the range of frequency in which this equipment operates, and that the term $(dC/dS)\Delta S$ gives the total required correction. Then Eq. (4) may be written

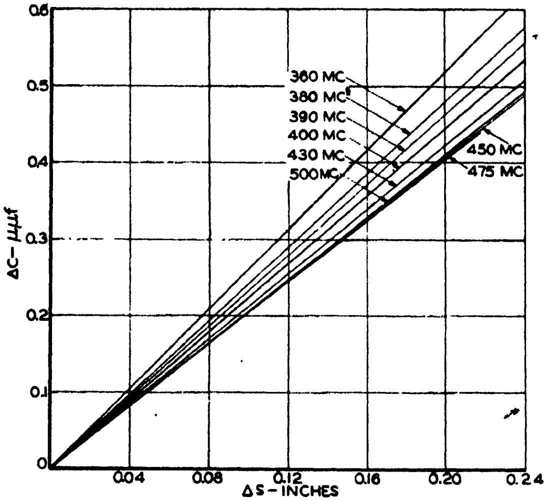


Fig. 8. Correction curve for the 400–600-Mc fixed cavity.

$$C_x=C_{a2}-\frac{dC}{dS}\Delta S. \qquad (5)$$

Replacing the term $-(dC/dS)\Delta S$ by its value, $\Delta C$, as obtained from the curves of Fig. 8 and writing for $C_{a2}$ its value for the case of electrodes 1″ in diameter, the following expression is obtained

$$C_x=\frac{0.1764}{\begin{array}{c}\text{Gap spacing (inches) at}\\ \text{resonance with sample}\\ \text{removed}\end{array}}+\Delta C\ \text{u.u.f.} \qquad (6)$$

The simplicity of Eqs. (3) and (6) permits very rapid calculation of the dielectric constant.

### 3. Expressions for the Dissipation Factor

An expression for the dissipation factor of a sample placed between the ends of the posts of a re-entrant cavity was derived in an earlier paper.[6] It is identical with the equation used for the "Change of Voltage Procedure" in the well-known susceptance variation method used at radiofrequencies.[10] The final result is given by the following expression:

$$\mathrm{Tan}\,\delta_x=(\Delta C_2/2C_x)(E_2-E_1)/E_1, \qquad (7)$$

where $E_1$ is the voltage at resonance with the sample in place, and $E_2$ is the voltage at resonance with the sample removed. The term $\Delta C_2/2C_x$ should be recognized as the dissipation factor of the empty cavity. $\Delta C_2$ is the change in capacitance necessary to tune from one half power level of resonance to the other.

## IV. PERFORMANCE

The dielectric constant and dissipation factor of a number of materials over a frequency range of 60 cycles per second to 10,000 megacycles per second are shown in Figs. 9 to 11. These curves are given to show the agreement in the results obtained using the instruments and techniques described with values obtained by independent methods. Symbols as listed on each curve are used to identify the method used to obtain a particular value.

Figure 9 shows the dielectric properties of polystyrene. The properties of this material are known to be practically constant over the frequency range shown on this curve. Therefore, the measurement of this material provides a

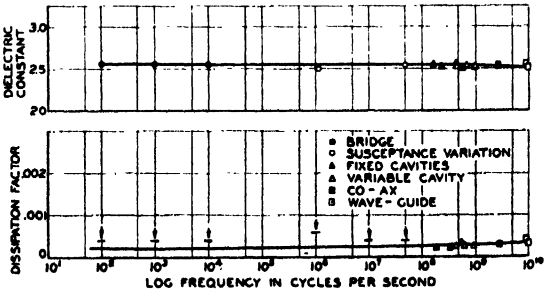


Fig. 9. Dielectric properties of polystyrene.

convenient check as to whether the measurements made by the method described agree with measurements made by independent means at higher and lower frequencies. Inspection of the curve for the dielectric constant shows that the values for the frequency decade $10^8$ to $10^9$ cycles per second lie along a curve already well defined by values determined by independent methods. Likewise, Fig. 10 shows that the values for the dielectric constant of insanol, which has a value of about 8, in the frequency decade in which we are interested do not depart from the curve. This shows that the accuracy does not vary with change in the value of the dielectric constant. The accuracy of the value of dielectric constant, as indicated by these curves, and other unreported measurements, is about ±1 percent.

The value of the dielectric constant of most solid materials is usually between two and ten. In the case of those unusual materials whose dielectric constant is considerably greater than ten, the samples should be much smaller than the electrodes.[17] A sample with a dielectric constant much greater than ten, which occupies the entire space between the electrodes, shortens the effective length of the cavity to such an extent that even when the physical length of the cavity has been made a minimum, it is not possible to resonate the cavity at frequencies in the upper portion of the band for which the instrument was designed.

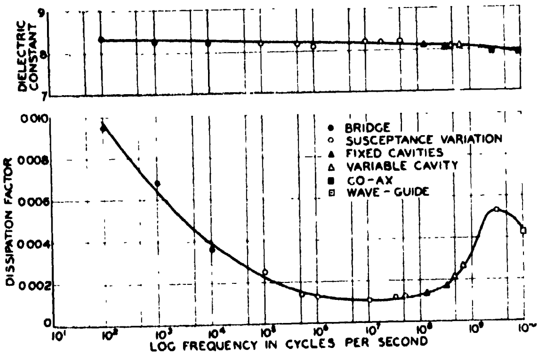


FIG. 10. Dielectric properties of insanol (mica-filled lead borate).

The error in the dielectric constant caused by variation in voltage across the face of the electrodes was investigated and the proper correction made in all cases where the error was not negligible. When materials having dielectric constants greater than five are measured at frequencies over 500 megacycles in the cavities described, a correction should be made for the variation in voltage across the electrode faces.[18]

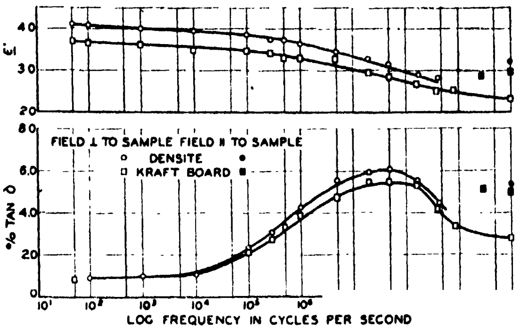


FIG. 11. Dielectric properties of kraft board dried at 105°C for 16 hours.

The curve for the dissipation factor, Fig. 9, shows that these values are in line with the values at higher frequencies obtained by wave guide measurements. At lower frequencies, the precision of the apparatus is ±0.0002 so that all that may be stated with certainty is that the dissipation factor is less than 0.0004. Further examination of the curves of dissipation factor, Figs. 9 and 10, shows that the accuracy of these values, as obtained by both the fixed and variable cavities, is ±0.00005 for small values of tan$\delta$.

In spite of the fact that certain approximations and empirical calibrations have been made to keep the measuring procedure and calculations nearly as simple as those of the conventional

---

[17] When the sample is smaller than the electrodes, Eq. (4) is applied to determine the over-all dielectric properties of the space between the electrodes (with the sample in place between the electrodes), and then standard equations for dielectrics in parallel (sample and air) are applied to determine the properties of the sample.

[18] The capacitance and, therefore, $e'$ of a sample placed between electrodes varies as: $e_0' = e_m'(E_{\max}/E_{\mathrm{Av}})$ where $e_c'$ is the corrected dielectric constant and $e_m'$ is the dielectric constant as calculated by Eq. (3). The average voltage is given by the expression $E_{\mathrm{Av}} = E_{\max}[1-(\pi^2 e' r^2 f^2/2c^2)]$ where $r$ is the radius of the sample, and the remaining symbols have been defined previously. This last equation may be derived by using Eqs. (7′12), (7′6), and (7′2) of reference 12, expanding $J_0$ in an infinite series, and taking the average. Because only the first two terms of the expression are given, $r$ must be less than $\lambda/8$, where $\lambda$ is the wavelength in air.

susceptance variation method, the accuracy of the values of the dielectric constant is very nearly the same as that obtained through the use of this technique at lower frequencies. The accuracy of the values of low dissipation factors is much greater than that obtained by any of the methods used at lower frequencies. This is due to the fact that very high values of $Q$ are obtained with the resonant cavity. This property of the resonant cavity may be taken advantage of through suitable techniques to obtain the dissipation factor of low loss dielectrics to very great precision.[19]

The curves of Fig. 11 are included because they illustrate the usefulness of apparatus operating in this frequency range, and point out some interesting properties of dielectrics. These materials have a dispersion region. The dielectric constant decreases as the dissipation factor goes through a maximum between $10^7$ and $10^8$ cycles per second. The apparatus described made possible an accurate determination of these curves throughout the dispersion region. Few laboratories have an apparatus for the measurement of solids operating in the frequency band of $10^8$ to $10^9$ cycles per second. These materials have another interesting property that should be mentioned in passing. The direction of the electric field in the sample makes a considerable difference in the values of the dielectric properties. When this material is placed in a wave guide with the same face against the end of the guide which was in contact with the electrode, the direction of the field in the sample is changed by 90°. Therefore, when changing from an instrument with electrodes to a wave guide, consideration should be given to the direction of the field in the sample in the case of materials which may be non-isotropic.

## V. CONCLUSION

Although the calibration of the apparatus described in this paper was accomplished by empirical methods, it has many advantages to recommend its use over methods which are mathematically more rigorous.

The sensitivity of these re-entrant cavities is so great that very small disk-shaped samples can be used. This makes it possible to sample directly the material to be measured. Thus, a piece of dielectric material may be removed from a cable or other apparatus and tested. When large samples of intricate shapes are required, as in the case of coaxial transmission line methods, especially at frequencies near the lower end of the frequency band (100–1000 Mc) considered in this paper, samples molded or cast in the laboratory may not be representative of a commercial grade of the same material. The simple disk shape of the samples required for this method makes their preparation easy.

Measurements may be made rapidly, and the calculation of results is extremely simple. The accuracy obtained from this method is usually better than that of other methods operating in this frequency range.

## VI. ACKNOWLEDGMENT

The author wishes to make grateful acknowledgment to Drs. L. J. Berberich and T. W. Dakin of the Westinghouse Research Laboratories, and to Dr. E. W. Montroll of the University of Pittsburgh for the many helpful suggestions given during the course of this work.

---

[19] William R. MacLean, "A microwave dielectric loss measuring technique," J. App. Phys. 17, 558–566 (1946).

# A Variable Capacitor for Measurements of Pressure and Mechanical Displacements; A Theoretical Analysis and Its Experimental Evaluation*

JOHN C. LILLY, VICTOR LEGALLAIS, AND RUTH CHERRY
*E. R. Johnson Foundation for Medical Physics, University of Pennsylvania, Philadelphia, Pennsylvania*

(Received December 13, 1946)

A variable capacitor is described for measuring (1) small displacements, (2) small volume changes, and (3) pressure differences. The capacitor consists of a deflectable diaphragm and a fixed electrode. The diaphragm is metallic, plane-parallel, clamped at the edges, and at ground potential; the electrode, at an a.c. potential, has a plane surface parallel to the undeflected plate across an air gap. For use in displacement measurements, the diaphragm's center is deflected by a point contact from a mechanical link to the observed system, or by a uniform pressure load from a fluid link to the system. The fluid link is used also when measuring volume changes and pressure differences. The plate deflection results in a change in the air gap, and thus generates a capacitance signal. This signal is measured by electrical methods.

A theoretical analysis of this variable capacitor is presented; sensitivity and alinearity factors for the three uses of the device are derived. The experimental performance shows reasonably satisfactory agreement with the derived theory. The displacement of the plate's center was measured with an interferometric method, using a yellow He line as a standard of reference; the applied pressure, with a liquid manometer; and the capacitance signal, with a standard capacitor substitution procedure. The gauge can be used so as to give quantitative electrical indications of displacement, volume change, or pressure difference; or can be used as a null indicator device in which an unknown pressure is balanced against a known one on opposite sides of the diaphragm. In order to achieve large volume an displacement sensitivities, small air gaps ($5.10^{-4}$ cm) are employed. Details of a construction method to assure small values are presented.

## INTRODUCTION

To anyone who has attempted the design or the construction of a diaphragm capacitance gauge for precise pressure or volume-change measurements, the lack of an adequate theory to handle the many variables encountered is rather discouraging. The important variables are the diaphragm diameter, thickness, and material; the air gap, the electrode diameter, the applied pressure, and the capacitance signal. The magnitude of the pressure signal and the desired capacitance signal are known. But what about the values to be assigned the first five variables in order to achieve the desired pressure-capacitance relation? It was an attempt to answer this question that led to this analysis.

Though the theory presented here was evolved in response to a need in the biological field, the treatment and the results are sufficiently general to be of interest to the physicist, the chemist, or the engineer who may want a gauge for other applications. For the benefit of those who may apply the gauge to biological systems, and for the benefit of those interested in why biological applications are as strict in their requirements as the physical ones, the peculiarities of the living systems are presented and discussed in relation to the capacitor gauge design. For those interested only in gauge design, this discussion on biological problems may be neglected, and a start made on the later, theoretical derivation section.

## THE GAUGE DESIGN PROBLEM IN BIOLOGICAL APPLICATIONS

The pressure gauge design problems encountered in the biological field are different from those in other fields of research. Certain limitations are placed on the pick-up unit design by the special properties of animals. In general, the pressure systems in animals are buried within the body, loosely supported, of variable distensibility, in constant motion, and are fragile in construction. These structures react locally to probing, puncturing, and tearing by bleeding and by constricting. Blood in the vessels may clot on inserted objects, and thus plug pressure taps. The whole animal also reacts by moving and by markedly changing the blood pressure, the heart rate, and other variables one may be measuring. Anesthetics can be used to abolish many of these undesired effects; but these drugs can exert other actions which are also unwanted.

---

* The work described in this paper was done under a contract recommended by the Committee on Medical Research, between the Office of Scientific Research and Development and the University of Pennsylvania.

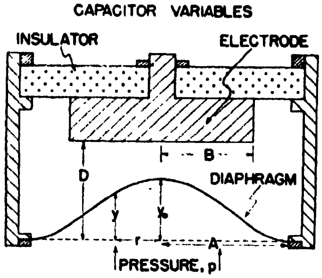


FIG. 1. Diagram of a diaphragm capacitor. The variables used in developing the capacitor theory are illustrated here; capital letters refer to fixed geometrical properties, small letters to variables generated by a displaced diaphragm. Tables I and II give the definitions and the units of all variables. The diaphragm is shown in an approach deflection; in a recession deflection the diaphragm would bulge downward. Both the air gap and the center deflection are greatly magnified in relation to the diameters of the diaphragm and of the electrode. The form of the deflected diaphragm is that of a stiff, edge-clamped plate; Figs. 2 and 3 give quantitative details of the two extreme idealized diaphragm forms assumed in the theory.

The pressure phenomena occurring in animals have both a static level and superimposed rapid pulsations. It is important to record these phenomena faithfully.

Because of these special properties of the animal, some form of transmission of the pressure signal from the interior to the exterior is necessary. At the present time, tubing containing liquid usually conducts the signal; ultimately, electrical transmission may be substituted. To avoid undesired local stimuli, the tubing is small.

From these considerations, some requirements for a pressure gauge pick-up can be deduced. (1) The pick-up must reproduce static level signals. (2) For the faithful reproduction of rapidly changing pressures transmitted through small bore tubing, and to minimize the length of the tube bore exposed to coagulable blood, the volume change of the unit with respect to an applied pressure must be minimal. (3) The pick-up size and the weight should be as small as possible. (4) The connection between the pick-up unit and the recorder must be long and very flexible. (5) The pick-up and its leads should be insensitive to vibration and to changes in position.

Many types of diaphragm pressure gauges have been used in recording static and kinetic pressures in animals.[1–8] The ones with the greatest promise of meeting the above requirements are the photoelectric[6, 7] and the electrical capacitor[8] types. The latter type is the one analyzed here.

With the above conditions in mind, it was considered important to examine the theoretical basis of the pick-up unit of a diaphragm-type capacitor pressure gauge previously applied to the recording of pressure in the vascular system.[8] The examination, presented below, was undertaken in order (1) to find out what factors were of importance in an optimal design for a pick-up unit for a minimal volume change and maximal capacitance signal in response to a given applied pressure; (2) to facilitate making design compromises in the presence of construction difficulties in special units; (3) to find and to evaluate those factors in the design which determine the alinearity and the sensitivity of the gauge; and (4) to explain and thus make allowances for the observed differences in gauge

TABLE I. Measured and calculated variables for a given gauge.

| I. Measured variables | Units | Method |
|---|---|---|
| 1. Electrode radius | $B$ cm | |
| 2. Diaphragm radius | $A$ cm | By micrometer |
| 3. Diaphragm thickness | $T$ cm | |
| 4. Applied pressure | $p$ dynes/cm² | By liquid manometer |
| 5. Central point deflection, at $p$ | $y_0$ cm | By interferometer |
| 6. Displaced volume, at $y_0$ | $\bar{v}$ cm³ | By calibrated capillary |
| 7. Gauge capacitance, at $y_0=0$ | $C_0$ cm | By substitution for standard condenser in an electrical circuit |
| 8. Capacitance signal, at $y_0$ | $c$ cm | |
| II. Calculated variables, as derived from measured ones | | |
| 9. Air gap, at $y_0=0$ | $D$ cm | $D=B^2/4C_0$ |
| 10. Displaced volume, at $y_0$ | $v$ cm³ | $v=(y_0)(\pi/k)A^2$ |
| 11. Relative deflection | $w$ (0) | $w=y_0/D$ |
| 12. Radius ratio, squared | $N$ (0) | $N=(B/A)^2$ |
| 13. Relative capacitance signal | $s$ (0) | $s=c/A^2/4D=(c/C_0)N$ |
| 14. Deflection constant, stiff | $K$ cm³/dyne | $K=(y_0/p)(T^3/A^4)$<br>$K=\frac{3}{16}(1-t^2)/Y$<br>($t$ = Poisson's ratio; $Y$ = Young's modulus) |
| 15. Deflection constant, flexible | $1/G$ cm/dyne | $1/G=(y_0/p)/A^2$ |
| 16. Volume constant, stiff | $k$ (0) | $k=y_0(\pi A^2)/v=3$ |
| 17. Volume constant, flexible | $k$ (0) | $k=y_0(\pi A^2)/v=2$ |
| 18. | $H$ (0) | $H=(1-N)$ |
| 19. | $u$ (0) | $u=w^2$ |
| III. Calculated variables, as related in theory: see equations in Table III | | |

---

[1] O. Franck, Zeits. f. Biol. **82**, 49 (1925).

[2] C. J. Wiggers, J. Lab. and Clin. Med. **10**, 54 (1924).

[3] W. F. Hamilton, Am. J. Physiol. **107**, 427 (1934).

[4] A. Hampel, Pflunger's Arch. Ges. Physiol. **244** (2), 171 (1940).

[5] W. G. Kubicek, Rev. Sci. Inst. **12**, 101 (1941).

[6] H. Rein, Arch. f. d. Ges. Physiol. **243**, 329 (1940).

[7] W. E. Gilson, Science **95**, 514 (1942).

[8] J. C. Lilly, Rev. Sci. Inst. **13**, 34 (1942).

[8a] F. Buchthal and E. Warburg, Acta Physiol. Scandinavica **55**, 55–70 (1943).

TABLE II. Sensitivity and alinearity factors for a constant $B/A$ and a given $w$.

| I. Sensitivity factors | Secant form | Tangent form | Initial value, $y_0 \to 0$ |
|---|---|---|---|
| 20. Relative displacement | $m = s/w$ | $n = (\partial s/\partial w)$ | $M = (1-H^k)/k$ |
| 21. Displacement | $c/y_0 = (m/4)(A/D)^2$ | $\partial c/\partial y_0 = (n/4)(A/D)^2$ | $S_y = (M/4)(A/D)^2$ |
| 22. Volume | $c/v = km/4\pi D^2$ | $\partial c/\partial v = kn/4\pi D^2$ | $S_v = kM/4\pi D^2$ |
| 23. Pressure | $c/p = mKA^6/4T^3D^2$ | $\partial c/\partial p = nKA^6/4T^3D^2$ | $S_p = MKA^6/4T^3D^2$ |
| 23a. Pressure | $c/p = (c/y_0)(y_0/p)$ | $\partial c/\partial p = (\partial c/\partial y_0)(\partial y_0/\partial p)$ | $S_p = (M/4)(A/D)^2(y_0/p)$ |
| II. Alinearity factors | | | |
| 24. Relative displacement | $a = m/M$ | $b = n/M$ | $a = b = 1.00$ |
| 25. Displacement | $a_d = (c/y_0)/S_y = a$ | $b_d = (\partial c/\partial y_0)/S_y = b$ | $a_d = b_d = 1.00$ |
| 26. Volume | $a_v = (c/v)/S_v = a$ | $b_v = (\partial c/\partial v)/S_v = b$ | $a_v = b_v = 1.00$ |
| 27. Pressure | $a_p = (c/p)/S_p = a$ | $b_p = (\partial c/\partial v)/S_p = b$ | $a_p = b_p = 1.00$ |

TABLE III. Theoretical equations relating $s$, $w$, and $B/A$.*

| | Stiff plate (Fig. 2) | | Flexible membrane (Fig. 3) |
|---|---|---|---|
| | Approach case | | |
| (2.0) | $s = (1/u)\tanh^{-1}[uN/(1-wH)] - N$ | (2.1) | $s = (1/w)\ln[(1-wH)/(1-w)] - N$ |
| (3.0) | $s = [(\frac{1}{3})(1-H^3)w + (\frac{1}{5})(1-H^5)w^2 + \cdots]$ | (3.1) | $s = [(\frac{1}{2})(1-H^2)w + (\frac{1}{3})(1-H^3)w^2 + \cdots]$ |
| (4.0) | $n = (\frac{1}{2})[1/(1-w) - H^3/(1-wH^2) - m]$ | (4.1) | $n = 1/(1-w) - H^2/(1-wH) - m$ |
| | Recession case | | |
| (5.0) | $s = (1/u)\tan^{-1}[uN/(1+wH)] - N$ | (5.1) | $s = (1/w)\ln[(1+w)/(1+wH)] - N$ |
| (6.0) | $s = [-(\frac{1}{3})(1-H^3)w + (\frac{1}{5})(1-H^5)w^2 - \cdots]$ | (6.1) | $s = [-(\frac{1}{2})(1-H^2)w + (\frac{1}{3})(1-H^3)w^3 - \cdots]$ |
| (7.0) | $n = (\frac{1}{2})[1/(1+w) - H^3/(1+wH^2) - m]$ | (7.1) | $n = 1/(1+w) - H^2/(1+wH) - m$ |

* Note that (Table I): $u^2 = w$; $N = (B/A)^2$; $H = (1-N)$.
The variables are defined in Figs. 1, 2, and 3, and in Table I. (1) Fundamental integral: $s = [(4D/A^2)\int_0^B (rdr)/2(D \pm y)] - N$. Solutions: [from (1) and equations of Figs. 2 and 3].

performance in the case in which the diaphragm is deflected away from the electrode in contrast to the case of diaphragm-electrode approach.

In addition to the measurement of pressure, a diaphragm-type capacitor has desirable characteristics for use as an ultramicrometer (displacement) and as a dilatometer (volume change). In the theory below, the condenser characteristics for these three uses are most easily discussed in the order: displacement, volume, and lastly, pressure. (In the theory, displacements are considered only as measured by means of a fluid link; in practice, point contact mechanical links can be used, but this theory does not apply in this case, even with a stiff plate.)

The following sections give a summary of a derivation of a diaphragm capacitor theory, and some applications of the theory in the design and in the testing of diaphragm capacitors. Later sections deal with the results of an experimental check of the theory, a consideration of electrical circuits for recording with the gauge, and a method used in the construction of some pick-up units.

## DERIVATION OF A THEORY OF A DIAPHRAGM-TYPE CONDENSER

### A. Configuration of the Capacitor

In deriving a theory of a diaphragm-type capacitor an expression is first developed which relates the displacement of the diaphragm to the resultant capacitance signal, for the geometrical arrangement shown in Fig. 1. In the following discussion, variables generated by a diaphragm deflection are represented by small letters, and the constants of a given gauge, by capital letters (see also Tables I and II).

The diaphragm is a circular plane-parallel metal plate or membrane clamped near the periphery and is assumed to be deflected by a uniform pressure load on one side or the other. The effective radius of this diaphragm, $A$, is that of the periphery at the inside edge of the clamp.

The electrode is a circular metal structure with a plane surface placed parallel to the undeflected plate, across an interposed air gap of a thickness $D$. The effective electrode radius, $B$,

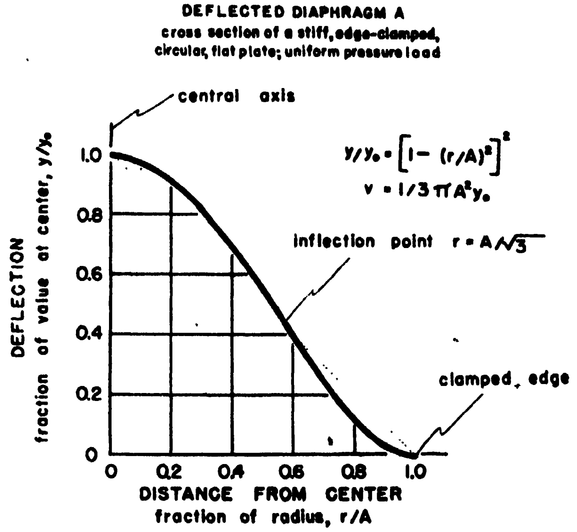


FIG. 2. The form taken by a stiff plate deflected by a uniform pressure. A radial section is cut through the diaphragm, from the axis of rotation at the left, to the clamped edge at the right. This method of showing the form greatly magnifies the amount of the deflection in relation to the radius by a factor of about one to ten thousand. The equation of this curve is derived on the assumptions that the absolute deflection is less than about 1/50 the thickness of the diaphragm, that the thickness is less than about 1/50 of the radius, and that the edge is uniformly and rigidly clamped, without exerting a radial tension in the undeflected position of the diaphragm (no initial stretching). All strains are assumed to be within the proportional elastic limit. It is further assumed that, in the undeflected state, the diaphragm is a plane, parallel, homogeneous, and isotropic structure. The equation of this curve is given in the figure; the variables are defined by Fig. 1. The volume swept out by the diaphragm as it is deflected, $v$, was found by integration of the equation for the curve; the resulting expression gives the dimensions of a cone of equivalent volume.

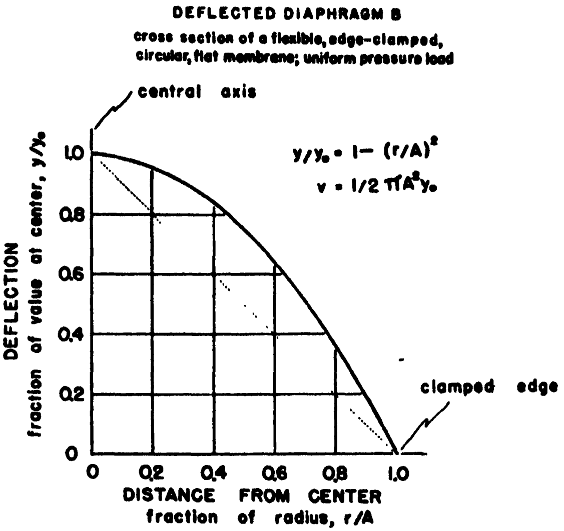


FIG. 3. The form taken by a flexible membrane deflected by a uniform pressure. The general statements made under Fig. 2 apply to this figure. The same limiting assumptions apply to this equation as to the stiff-plate one, except that it is assumed that the membrane is subjected to a uniform tension of $G/4$ cm/dyne (item 15, Table II) in all directions in its own plane in the undeflected state. In this diaphragm the inherent stiffness is assumed to be negligible, and hence the clamped periphery bends with a practically infinitesimal radius of curvature. The volume swept out by the diaphragm as it is deflected, $v$, equals that of a cylinder of a height equal to one-half the central deflection and a base of an area equal to that of the undeflected diaphragm.

is that of the plane surface presented to the air gap.

The diaphragm and the electrode have a common axis passing through their centers, normal to their facing surfaces (Fig. 1).

Since the ratio of the electrode periphery to the air-gap thickness, $2\pi B/D$, is usually of the order of magnitude of 750–7500; edge effects are considered to be negligible and are omitted in the following analysis.

### B. Derivation Summary

An equation defining the relation of the relative change in capacitance of this capacitor to the deflection of the diaphragm (Fig. 1) can be derived in the following manner:

1. Consider an elemental capacitor consisting of an annulus of infinitesimal width, lying in the surface of the electrode and concentric with the central axis, and of a similar strip directly across the air gap, lying in the surface of the diaphragm. Let both strips have a radius of $r$, defined as the distance along a normal from the annulus to the axis. Assume that, as the plate is moved by a pressure load, the capacitance of this elemental capacitor, $dc$, (in centimeter units) varies inversely with the distance between the strips, and varies directly with the area of one strip ($2\pi r dr$) (Eq. (1), Table III, note).

In the usual case, the deflection of the diaphragm center is only about 1/500 to 1/5000 of the diaphragm radius; therefore it can be shown that the above assumption is justified.

2. Consider the annular element in the surface of the diaphragm in the deflected position. The displacement of this element from the undeflected position is taken as $y$, and that of the diaphragm center as $y_0$. The relation between the radius of the annulus, $r$, and $y$, should be con-

sidered for at least two extreme cases, the thin, stiff plate and the thin, flexible membrane.[9]

The equation relating $y$, $y_0$, $r$, and $A$ for the stiff plate is given in Fig. 2; and that for the flexible membrane in Fig. 3. (These equations are taken from Love's *The Mathematical Theory of Elasticity*, and reference 9.) The equations are exact only for values of $y_0/A$ which are very small compared with unity. Other assumptions and limits for these equations are given in the Figs. 2 and 3 captions.

3. By setting up the differential equation implied in assumption 1 above, substituting the expression for assumption 2, for the two diaphragm types and integrating the result, the equations of Table III can be derived. These equations, in Table III, state the relation between the relative capacitance signal, $s$, and the relative change in the air gap, $w$ (equals $y_0/D$). In these expressions the parameter is the electrode-diaphragm radius ratio, $B/A$. All of the deflections of the diaphragm toward the electrode are treated as the "approach case," and deflections away from the electrode as the "recession case.[10] The equations of these two cases differ fundamentally only in the algebraic sign of the central diaphragm displacement term, $y_0$, which is positive in recession and negative in approach.

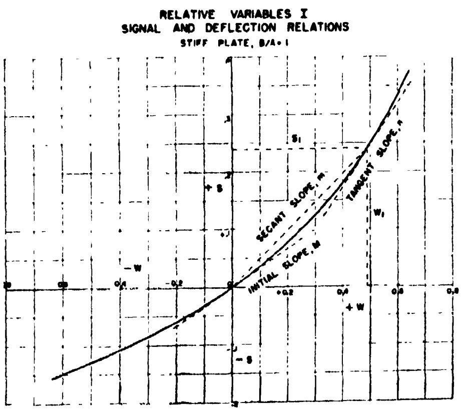


Fig. 4. Relation between the relative displacement and the relative capacitance change for a stiff plate. This curve is for an electrode and a stiff plate diaphragm (Fig. 2) of equal radii. For any condenser with a given radius ratio, $B/A$ (Fig. 1), the fractional, or relative, capacitance change, $s$, is defined as the capacitance change, $c$, divided by the undeflected state capacitance, $C_0$, of a condenser with a $B/A$ of unity and the same initial air gap, $D$ (Fig. 1 and item 13, Table I). The fractional air-gap change or the relative displacement, $w$, is defined as the central diaphragm deflection, $y_0$, divided by the air gap, $D$, in the undeflected state (Fig. 1). This figure serves to define the relative displacement sensitivities, $M$, $m$, and $n$ in terms of the slopes of the $s$ and $w$ curve. The secant $m$, and the tangent $n$, are both defined at any point on the curve, $s_1$ and $w_1$; the initial slope, $M$, is that at the origin, $s=w=0$. The sole dependence of the initial slope, $M$, on the radius ratio, $B/A$, is shown in the next figures. Both $m$ and $n$ approach $M$ as the deflection, $w$, becomes infinitesimal. For the assumptions and the theory of this curve see text. It is to be noted that the variables used in this and in most of the rest of the figures have zero dimensions.

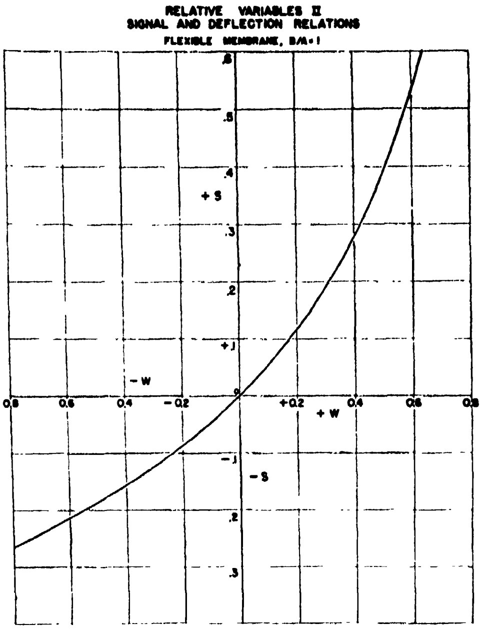


Fig. 5. Relation between the relative displacement and the relative capacitance change for a flexible membrane. The figure is the same as Fig. 4 with the exception that the diaphragm here is a flexible membrane (Fig. 3) instead of a stiff plate. The definitions of the variables are given under Fig. 4; the theory is developed in the text.

For calculation purposes, the arc tangent, the arc hyperbolic tangent, and the logarithmic forms of the solution were used for numerical values of $y_0/D$ greater than 0.10, and the series forms for the lesser values (0.00 to 0.10).

## USE OF THE THEORY IN THE DESIGN OF PICK-UP UNITS

### A. Factors Influencing the Sensitivity

In general the final objective in a problem of the design of a pick-up unit is reached when the unit produces a capacitance signal of the mag-

---

[9] I. B. Crandall, *Theory of Vibrating Systems and Sound* (D. Van Nostrand Company, New York, 1927).

[10] Roess gives a stiff plate approach case equation which is equivalent to the inverse hyperbolic tangent function in Table III. [L. C. Roess, Rev. Sci. Inst. 11, 183 (1940).]

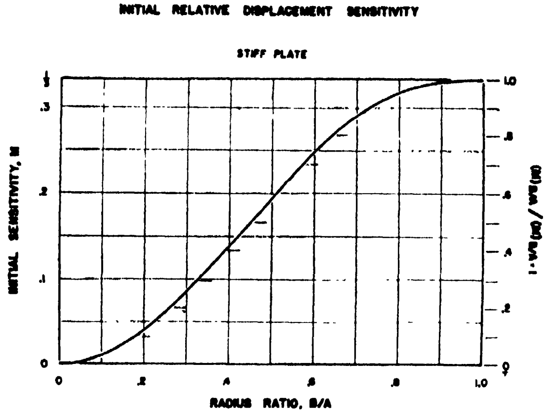


FIG. 6. The dependence of the initial relative displacement sensitivity on the radius ratio for a stiff plate. As is shown in the theory in the text, the initial relative displacement sensitivity, $M$ (Fig. 4), is a function solely of the radius ratio, $B/A$ (Fig. 1). The right-hand scale shows the sensitivity at a given $B/A$ as a fraction of that at the maximum value, at $B/A$ of unity. The initial absolute sensitivity, defined as the absolute capacitance change, $c$, divided by the absolute central deflection, $y_0$ (Fig. 1), can be calculated from these values of $M$, the electrode radius $B$, the diaphragm radius $A$, and the initial air gap, $D$ (see text). The ratios of the secant, $m$, and the tangent, $n$, relative sensitivities (Fig. 4) to $M$ for a number of radius ratios is shown in later figures.

nitude required by a chosen electrical circuit in response to an applied displacement, volume, or pressure change of a size determined by the system in which measurements are to be made. In other words, the designer's problem is to achieve a definite gauge sensitivity. In most biological applications to pressure recording, as was shown above, an additional part of the final objective is to have the volume of the gauge change a minimal amount. In some applications, a minimal alinearity of the calibration curve is also desirable.

Two useful types of sensitivity factors can be defined for diaphragm gauges; the absolute and the relative types. The absolute factors are defined in terms of the variables measured while the gauge is in use, which are the displacement, the volume, or the applied pressure, and the resulting capacitance signal. The relative sensitivity factors are expressed in terms of dimensionless quantities, such as those used in the theoretical equations developed above. After the gauge constants are known, the absolute sensitivity factors can be calculated from the relative sensitivity ones. This calculation will be dealt with in a later section.

## B. Relative Sensitivity

Figures 4 and 5 show the graphs of the function relating the relative capacitance signal, $s$, to the relative displacement factor, $y_0/D$, for a radius ratio, $B/A$, of 1.00 (Tables I and III). As the plate or membrane deflects an infinitesimal amount in either direction ($y_0/D$ approaching zero), the slope of this curve has a definite value. From inspection of the series forms of the equations in Table III, it can be seen that this slope at the origin is a function only of the radius ratio, $B/A$; this slope is called the "initial relative displacement sensitivity," $M$, for a given value of the radius ratio, $B/A$. Figures 6 and 7 show the relation between $M$ and $B/A$ for the two extreme diaphragm types.

As the relative displacement, $w$ (equal to $y_0/D$, Table I), increases beyond an infinitesimal value, the relative sensitivity changes from the initial value, $M$. Two useful types of relative sensitivity factor can be defined for the rest of the curve; these two types are illustrated in Fig. 4. The slope of a straight line from any point on the

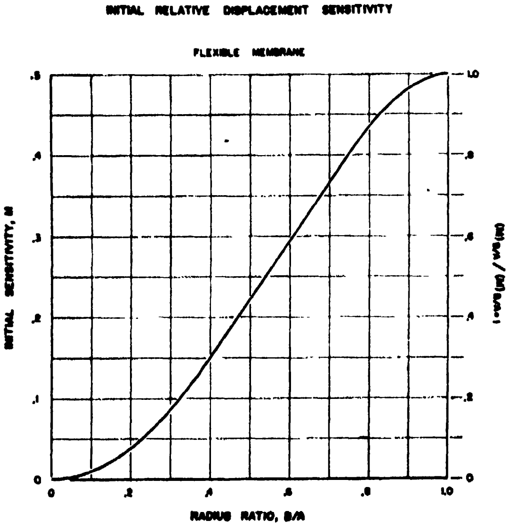


FIG. 7. The initial relative displacement sensitivity and the radius ratio for a flexible membrane. The statements under Fig. 6 apply also to this figure. It is to be noted that the maximum value of $M$ at a $B/A$ of unity for the flexible membrane is one-half, and that for the stiff plate (Fig. 6) is one-third. From the expression for the initial absolute volume sensitivity, $S_v$, (item 22, Table II), and these values of $M$ and the volume constant, $k$ (16 and 17, Table I), it can be shown that the initial absolute volume sensitivity for the stiff plate is equal to that of the flexible membrane considering all other geometrical values as identical, and $B/A$ as unity.

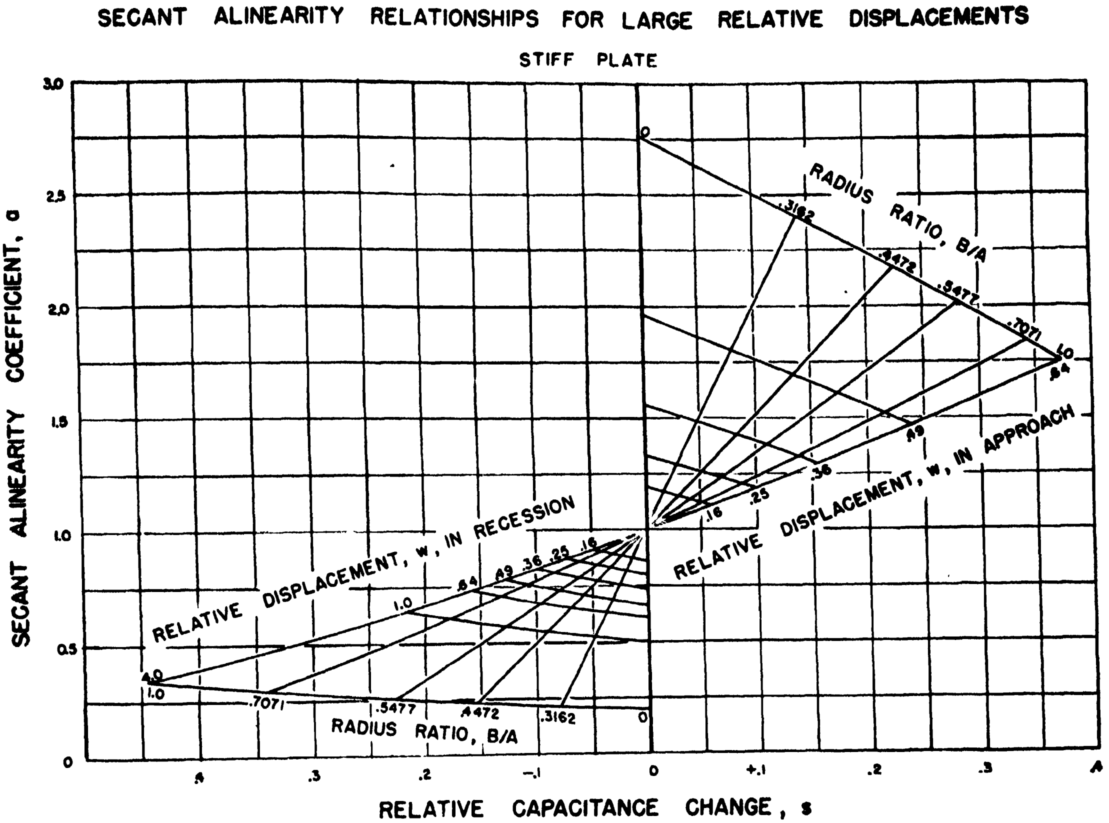


FIG. 8. The secant alinearity coefficient and the relative capacitance signal relations for a stiff plate for large relative deflections. Some of the variables used in this graph are defined under Fig. 4, all variables are defined in Tables I and II. The secant alinearity coefficient, $a$, is the ratio of the secant relative displacement sensitivity, $m$, to the initial displacement sensitivity, $M$. This figure shows the dependence of $a$ on $B/A$, $s$, and $w$ for a stiff plate. A point to be noted is the large deviations from linearity in the approach condition as contrasted with the smaller deviations in recession, for corresponding values of $w$.

curve to the origin is called the "secant relative sensitivity," $m$, which is equal to $s/w$, Table I.

The slope of the curve itself at the same point is the "tangent relative sensitivity," $n$, which is equal to $\partial s/\partial w$. Both $m$ and $n$ approach the initial relative sensitivity value, $M$, as the relative displacement approaches zero, this property of these functions provides a convenient definition for two useful alinearity coefficients.

## C. Gauge Alinearity

Two alinearity coefficients, $a$ and $b$, can be derived from the two displacement sensitivity factors, $m$ and $n$. The "secant alinearity coefficient," $a$, is defined as the ratio of the secant sensitivity to the initial sensitivity, $m/M$, and the "tangent alinearity coefficient," $b$, as $n/M$. Both of these coefficients are defined for a constant value of the radius ratio, $B/A$, and are a function of $w$.

These two coefficients have the value of 1.00 as the relative displacement of the diaphragm approaches zero, i.e., a capacitor gauge is most linear for those diaphragm deflection values which are a very small fraction of the air gap value. This relation is true for the opposite cases of recession and approach, for both types of diaphragm; in recession, the coefficient values are less than 1.00; in approach, greater than 1.00. Figures 8 through 13 show these relations between $a$, $b$, $s$, $w$, and $B/A$ for the two diaphragms.

If the displacement of the center of the diaphragm, $y_0$, is a linear function of the volume change, $v$, or of the applied pressure, $p$, both coefficients give the alinearity values for actual

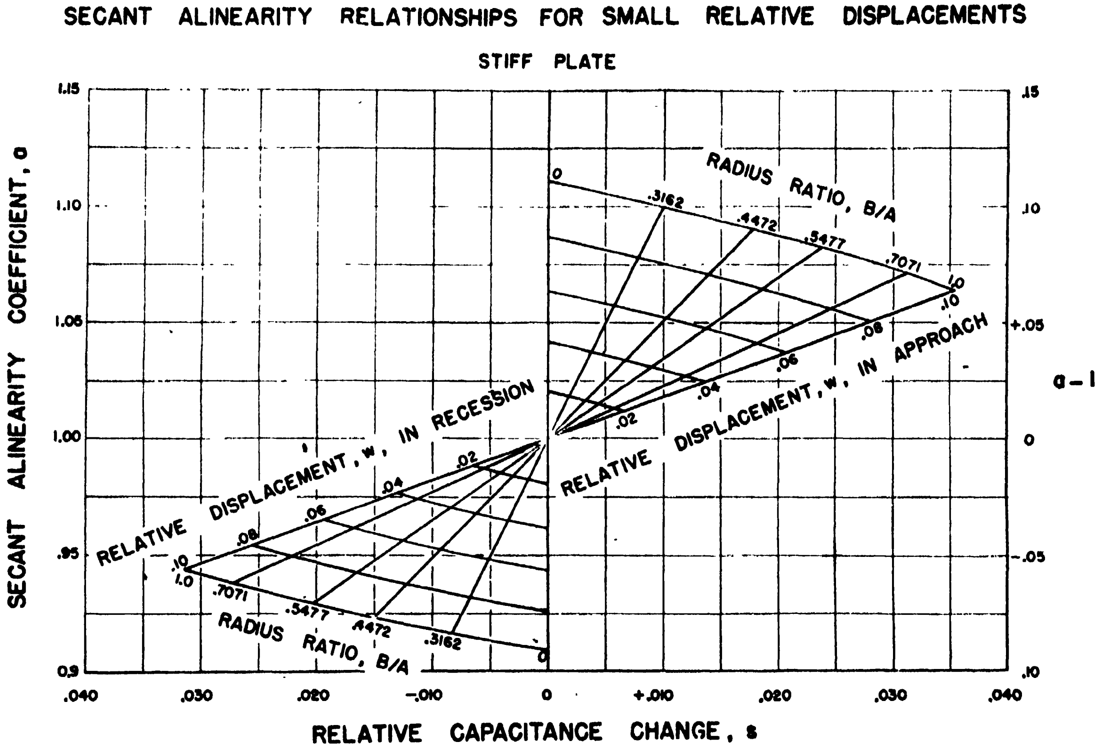


FIG. 9. The secant alinearity coefficient relationships for a stiff plate for small values of the relative displacement. This figure is an amplification of the part of Fig. 8 near zero capacitance change. In those uses of the gauge requiring a high degree of linearity, the variables are restricted to the range on this plot or to an even smaller range. The considerations mentioned under Fig. 8 apply also to this figure.

volume and pressure gauges. In other words, for a constructed gauge the ratio of the slope of the volume change-capacitance curve, or of the pressure-capacitance curve at a given pressure or volume, to the initial slope is given directly by the value of the alinearity coefficients at a given $s$, the relative capacity signal for a given radius ratio, $B/A$. A mechanical design which has a fairly linear pressure-deflection relation is discussed in the section on experimental performance.

### D. Absolute Types of Sensitivity

The absolute sensitivity factors are divided into the two general types, secant and tangent, defined above for the relative displacement factors.

It is convenient to define each of the secant factors as the ratio of the capacitance signal, $c$, to the displacement, $y_0$, to the volume, $v$, or to the pressure, $p$. The tangent sensitivities are defined as the rate of change of the capacitance with respect to each of the three variables. All of these absolute factors are calculable from the gauge constants, $A$, $B$, $D$ (or $C_0$) and the corresponding relative displacement sensitivities, $m$ and $n$; however, it may be easier to use these constants, the initial relative displacement sensitivity, $M$, and the two alinearity coefficients, $a$ and $b$. The equations for these calculations are given in Table II.

The absolute displacement sensitivity, $(c/y_0)$, for a given value of $B/A$, can be shown to be equal to one-fourth the product of the square of the ratio of the diaphragm radius to the air gap, $(A/D)^2$, and the relative sensitivity, $M$ (Table II); $c/y_0$ has an initial value of $(c/y_0)_0$ when $m$ equals $M$. The corresponding tangent factor, $dc/dy_0$, is similarly calculable by substitution of $n$ for $m$ (Table II). The values of $a$ for these and later calculations for various values of $B/A$ and $w$ are given in Figs. 8, 9, and 10; the $b$ values, in Figs. 11, 12, and 13.

In other words, the absolute displacement

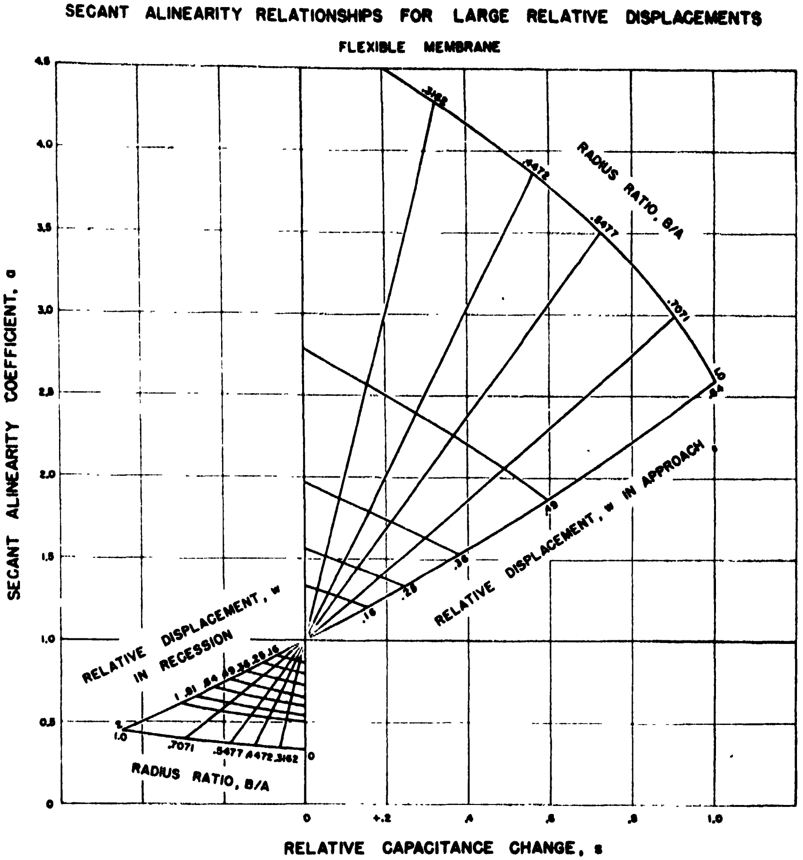


FIG. 10. The secant alinearity relationships for a flexible membrane. This plot for a flexible membrane is of the same type as the one for the stiff plate shown in Fig. 8. Comparing this figure with number 8 it can be seen that the flexible membrane has greater deviations from ideal linearity ($a$ equal to 1.0) than does the stiff plate for the same value of the relative displacement, $w$.

sensitivities, secant and tangent, are independent of the absolute values of the diaphragm radius, thickness, and material; of the electrode radius; and of the air gap. However, these factors are dependent on the ratio of the diaphragm radius to the air gap, and on the relative displacement sensitivity.

The absolute volume sensitivity factors, for a given value of $B/A$ are calculable by the equations given in item 22, Table II; or as in the case of $c/y_0$, by the substitution of $M$ and $a$ or $b$, for $m$ or $n$ in the equations. By these equations it can be shown that the two absolute volume sensitivities for both diaphragm types are independent of the absolute value of the diaphragm radius, thickness, and material; and of the electrode radius. However, the volume sensitivity varies inversely with the absolute value of the initial air gap. This result suggests that the air gap be designed and constructed with as small a value as is practicable in those gauge applications in which the volume change is desired minimal.

There are two alternative ways of calculating the absolute pressure sensitivities; the equations are given as items 23 and 23a in Table II. The first method (item 23) in the stiff-plate case involves knowing at least two of the elastic constants of the material of the diaphragm (Table I, item 14; Young's modulus and Poisson's ratio) with sufficient accuracy for the design problem, and using these to calculate the deflection constant.[11] In the case of the flexible membrane, it involves knowing the value of the initial stretch tension ($G/4$, item 15, Table I) in the membrane to calculate the deflection constant.[9] It was quickly found that the literature lacks the data necessary for a highly accurate calculation of the deflection constant for a stiff plate; and that the methods of measuring or calculating the initial stretch were unsatisfactory in the case of the flexible membrane. However,

---

[11] R. J. Roark, *Formulas for Stress and Strain* (McGraw-Hill Book Company, Inc., New York, 1938), p. 171, Case 6.

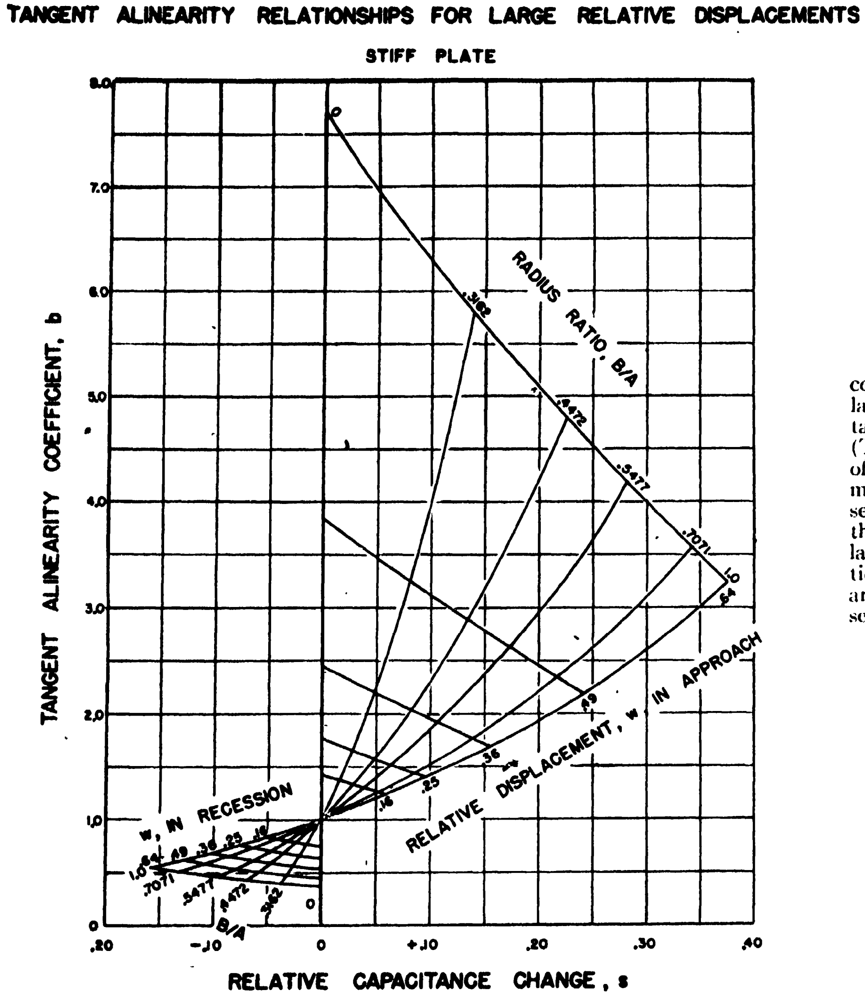


FIG. 11. The tangent alinearity coefficient for a stiff plate for large relative displacements. The tangent alinearity coefficient, $b$, (Table II), is defined as the ratio of the tangent relative displacement sensitivity, $n$, to the initial sensitivity $M$. A comparison of this plot with Fig. 8 shows the large deviations of the variational sensitivity ($n$) from linearity compared with the secant sensitivity, $m$.

for a first approximation in a design problem, this first method is useful.

The second method (item 23a, Table II) of calculating the pressure sensitivity envolves an independent measurement of $y_0/p$ with a diaphragm of the desired material, thickness, and radius. Figure 14 shows such a measurement on a stainless steel diaphragm by an interferometer method. With the determined value of $y_0/p$, the deflection constant can be calculated (Table I) and used in further design modifications to obtain any desired $c/p$ in an actual gauge. The data of Fig. 14 were used to calculate theoretical values of $c/p$ in the evaluation of this design theory (Fig. 15).

## EVALUATION OF THE DESIGN THEORY

Using a stiff-plate gauge shown in Figs. 16 through 19, data were taken on the pressure-capacitance performance ($c/p$ ratios) in the approach and recession conditions (Fig. 15). To check these data against the theory, theoretical $c/p$ values were calculated from the equation in item 23a (Table II) using the measured values of the electrode radius, $B$, the diaphragm radius, $A$, the initial capacitance, $C_0$, the displacement-pressure ratio, $y_0/p$, the value of $M$ from Fig. 6, and the values of $a$ from Figs. 8 and 9. The values of the $y_0/p$ ratio were determined by an interferometric method (Fig. 14), rather than by the less accurate method of calculation from the elastic constants. The results of these calculations are plotted on the graph of Fig. 15 with the experimental data. Figure 15 shows that the stiff-plate theory can be used with a fair degree of accuracy for design purposes, even in the

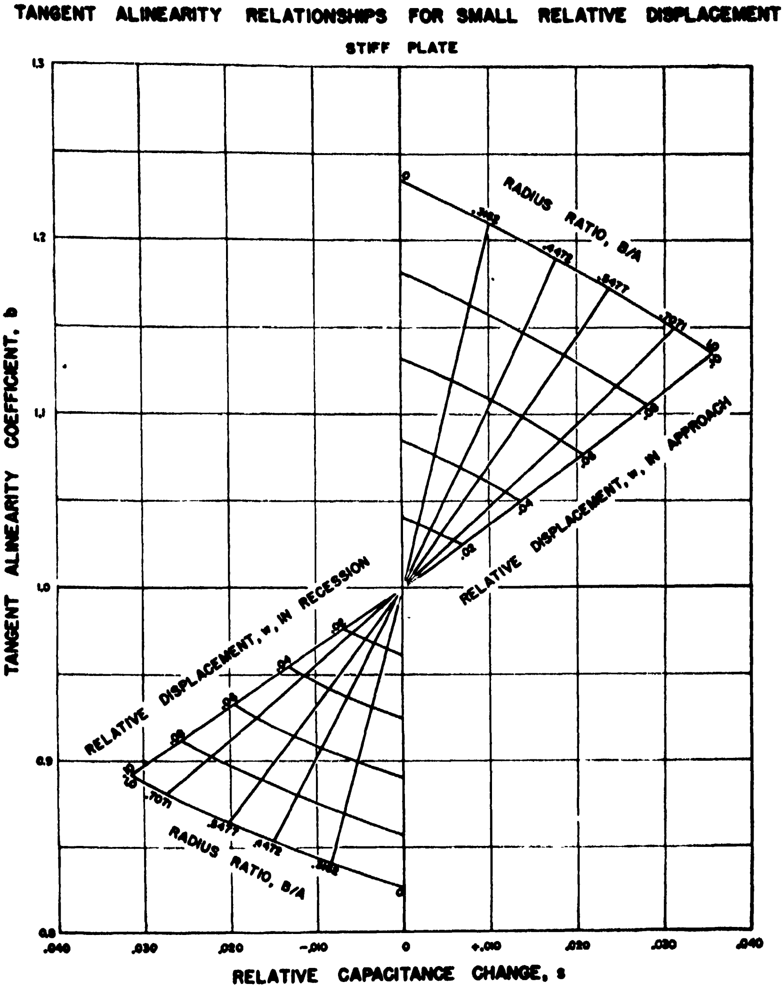


FIG. 12. The tangent alinearity coefficient for a stiff plate for small relative displacements. This plot is an enlarged view of Fig. 11 near the relative displacement of zero magnitude. This is a companion figure to number 9 for designs involving a high degree of linearity.

extreme cases involving large relative displacements ($w = \pm 0.49$) in approach and in recession. The flexible membrane theory still awaits confirmation; it is hoped that this can be done and reported in the near future; there are a few unpublished data which show that reasonably accurate predictions of performance can be expected from the flexible membrane theory.

## SPEED OF RESPONSE OF A CONSTRUCTED GAUGE

Figure 20 shows the response of a gauge with an attached needle to a sudden drop in pressure level. Details of the technique of obtaining the records is given under the figure; the gauge itself is pictured in Figs. 17, 18, and 19. These records show the importance of having a homogeneous conducting medium between the source and the gauge diaphragm. They also show that fairly high speed pressure recording can be accomplished through small conduction tubing with a relatively incompressible fluid, if the volume change of the gauge is small and if the tubing wall is relatively stiff. The volume change (Table I) for the full pressure range in Fig. 20 is approximately $2 \times 10^{-5}$ cc. The contained liquid volume is 0.4 cc; therefore an additional 20 percent must be added to the gauge volume change to allow for the compression of the contained water at a pressure of 150-mm Hg above 1 atmosphere.

For the speed of response of the gauge without an attached fluid system it was found that it was difficult to obtain a step change in pressure level sufficiently fast to approach the lowest natural mode of vibration of the stiff-plate diaphragms used (10 to 60 kilocycles). However, the applica-

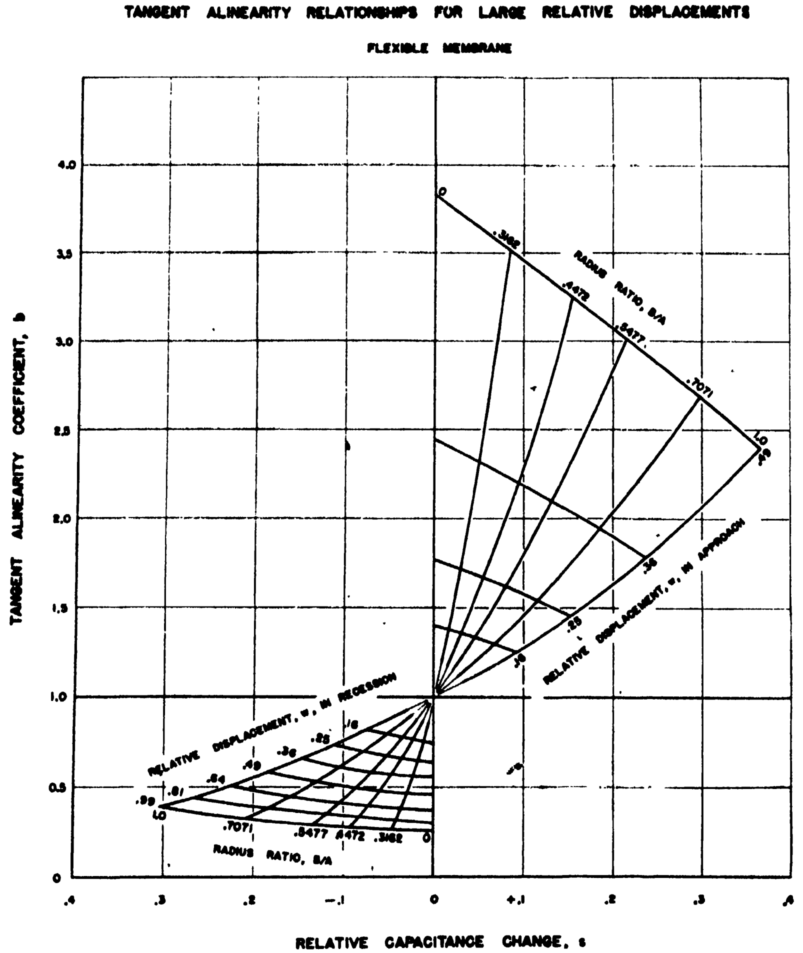


FIG. 13. The tangent alinearity coefficient for a flexible membrane. This figure is to be compared to Fig. 11 for a stiff plate. The tangent alinearity coefficients of the two cases do not differ to a great degree, in contrast to the secant alinearity coefficients (Figs. 8 and 10).

tions of the gauge to date have not made full use of the available speed of response.

## CONSTRUCTION DETAILS TO OBTAIN SMALL AIR GAPS

Standard machining is performed on all components (Fig. 16) except those having critical surfaces which require special techniques as follows:

In the electrode barrel assembly, the electrode (2, Fig. 16) and its lava insulators (6) are assembled to the barrel (4) with bakelite laquer, riveted (12, Fig. 17) and baked. This assembly is faced in the lathe and hand-lapped on a plate-glass lap, until the face of both the electrode and its barrel are in the same flat plane. The electrode spacing is generated by hand-lapping the electrode face with a brass lap. The amount of spacing thus generated is gauged by placing the assembly against a flat metal gauge block and measuring the electrical capacitance between the electrode and the gauge block. If the desired spacing is exceeded, correction may be made by lapping the face of the barrel on the plate glass.

For a given diameter, diaphragms are handled differently as they vary in thickness. A diaphragm of $\frac{11}{16}''$ outside diameter could be classified thick if it were 0.008″ and over, medium 0.004″ to 0.008″, thin 0.004″ and under.

Thick diaphragms are machined from bar stock and hand-lapped to size on plate glass. Frequent turning of the diaphragm, exposing first one face then the other to the lap, will minimize warping from strain release. The thickness of the diaphragm is gauged with a micrometer. The flatness is checked by optical means or by substituting the diaphragm for the gauge block in the capacitance test described above.

Medium diaphragms are disks cut from polished, hard rolled strip metal of the required thickness. The disks are generally curved in a cylindrical plane but approach a flat surface when clamped in position.

Thin diaphragms are disks cut from polished, half-hard rolled metal of the desired thickness. They require stretching while clamping to present flat surfaces. The stretching mechanism (not shown) consists of a recess in the electrode barrel face into which the diaphragm is pressed by a ring member. The ring member is designed with clearance to insure positive seating of the diaphragm.

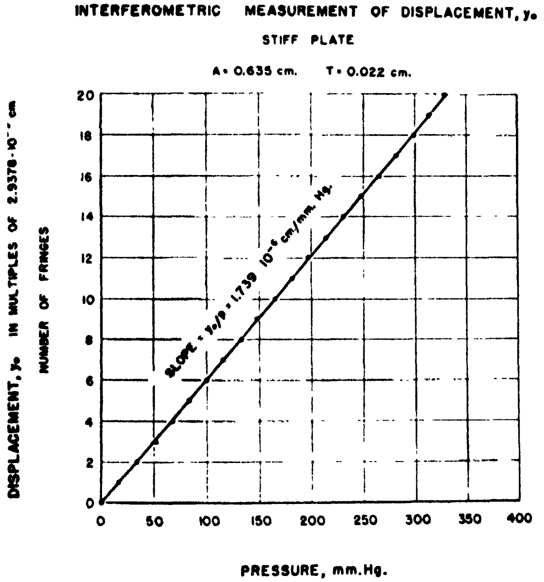


FIG. 14. Experimental determination of the pressure-displacement relation for a stiff plate. The plate is edge-clamped, and corresponds to the diaphragm of the type shown in Fig. 2. The capacitance-pressure relations for this unit are presented in Fig. 15, and the unit itself in Figs. 17, 18, 19, and 20. The absolute central diaphragm displacement, $y_0$ (Table I) was measured with a vertical incidence Newton's fringe type of interferometer, using a yellow line (5875.6 Angstroms) of a helium discharge tube. The fully aluminized plate (1×3×3 mm) of the interferometer was moved toward the 60 percent aluminized fixed plate by displacement of the diaphragm's center by an applied pressure. The pressure was measured with a compensated mercury manometer with an accuracy of ±0.1 mm Hg. The fringes were counted by means of a microscope and an eyepiece scale. Each fringe counted represents a movement of the diaphragm center of one-half of a wave-length of the He line, given on the left of the figure. The pressure range usually used with this gauge unit is that represented by the first two fringes, or 30 mm Hg. The diaphragm displacement is away from the electrode, i.e., in the recession condition.

The strict proportionality between the displacement and the applied pressure up to at least 300 mm Hg (ten times the range actually used) shows that this diaphragm is operated well inside of the proportional elastic limit.

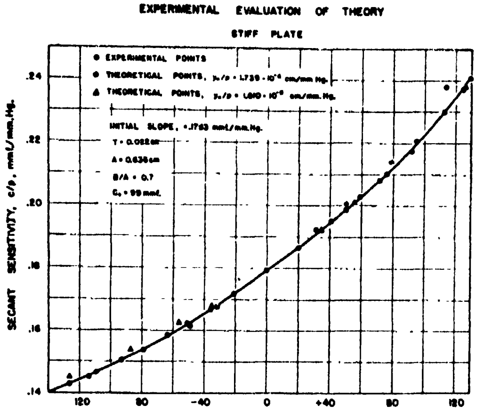


FIG. 15. An experimental evaluation of the design theory for a stiff plate. The capacitance change, $c$, and the initial capacitance, $C_0$, were measured with a "$Q$" meter at a frequency of 3.0 megacycles and by substitution of a precision condenser (General Radio Company). The pressure was measured with the Hg manometer mentioned under Fig. 14. The absolute pressure secant sensitivity factor, $c/p$, was calculated from the measured values of $c$ and $p$; the solid dots and the drawn curve represent these experimental data. The dotted circles are theoretical points calculated from the value of the recession state $y_0/p$ given in Fig. 14, the $B/A$ value of 0.7 (measured $B$ and $A$), the $C_0$ value of 99 mmf, the $M$ value of Fig. 6, and the secant alinearity coefficient values of Fig. 8, by means of the equation given as item 23a in Table II. It can be seen that in recession these points fit the experimental curve fairly well, but deviate from the curve in the high pressure region in the approach state. It was then found, by the interferometric method of Fig. 14, that the value of $y_0/p$ in the approach direction was smaller than in recession (value given in this figure). The points calculated with this value of $y_0/p$ are the triangles. Presumably the difference of $y_0/p$ in the two directions is due to unsymmetrical clamping supports on the two sides of the plate.

The main body (3, Fig. 16) has an interior seat against which the diaphragm is clamped. This seat is hand-lapped to a flat surface with a formed brass lap. The exterior faces of the main body are hand-lapped on plate glass to present flat seats to the mating attachments (13 and 15, Fig. 17), which are similarly finished.

All lapping operations referred to above were performed with a fine grain abrasive and light oil. Silicon carbide No. 400B was used for coarse grinding and No. 600B for finish.

## ELECTRICAL CIRCUITS FOR USE WITH THE GAUGE

The circuit actually used with these gauges will be published later. The choice of a given

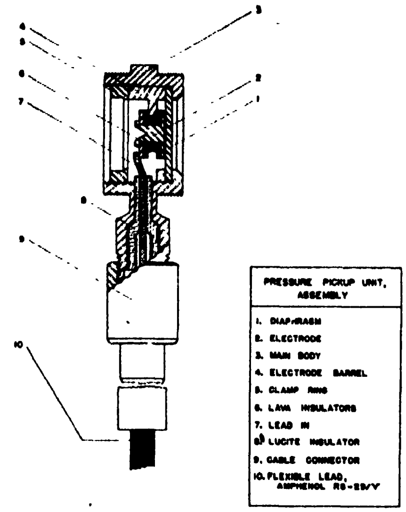


FIG. 16. Sectional drawing of a variable capacitor. The active diaphragm (1) diameter is 0.50 inches. The radius ratio $B/A$ is 0.70. The electrode-diaphragm air gap is $5.10^{-4}$ cm. (0.0002 inch). The pertinent parts are labeled in the figure. The lead-in (7) carries the r.f. voltage to the electrode; the rest of this unit is grounded. It is to be noted that the electrode side of the diaphragm is kept as open as possible to allow quick venting of the gas in this space in applications in which large ambient pressure changes occur, such as in the explosive decompression of pressurized aircraft cabins in altitude chambers. The lava insulators (6) are supported by three equally spaced legs protruding inward radially from the electrode barrel (5); only one of these legs shows in this section. This construction allows the quick venting mentioned above. The construction details to assure small air gaps are given in the text.

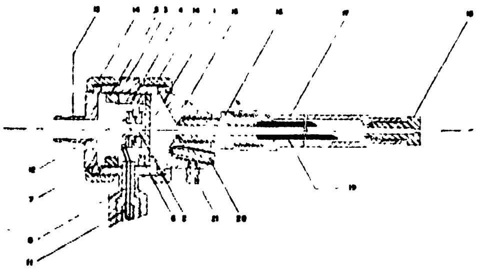


FIG. 17. Sectional drawing of a capacitor with an attached liquid system. Through number 8, the parts are the corresponding ones of Fig. 17. Photographs of this unit are shown in Figs. 18 and 19. A tubing attachment (13) for pressure calibration purposes is shown fastened to the electrode side of the unit by the retaining ring (14). A liquid system (15) with an attached hypodermic needle (16 and 19) is fastened to the diaphragm side of the unit. For the sake of clear reproduction the needle is greatly enlarged; the true relative size is shown in Fig. 18; a side tube (21) and a needle valve (20) are used for filling the liquid system. A needle guard (17) and its terminal plug (18) keep the needle wet and sterile until ready for use. This type of unit has been used for blood pressure, intrathoracic pressure and intraspinal pressure recording (Figs. 21 through 24); its response is shown in Fig. 20.

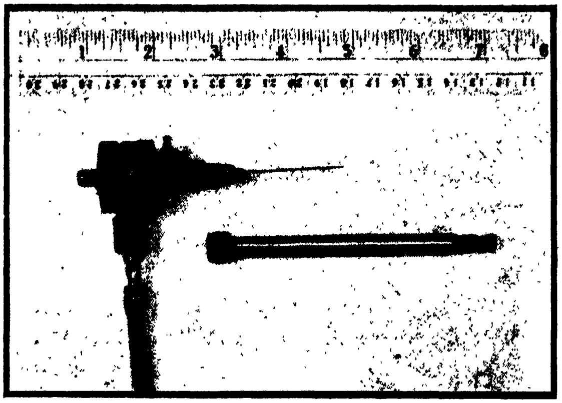


FIG. 18. Internal body pressure condenser gauge. A cross section drawing of this unit and its details is shown in Fig. 17. A photograph of this unit and its attachments is given in Fig. 19. The needle guard is below the needle. This unit is used for blood pressure and other internal body pressure recording (Figs. 23 and 24).

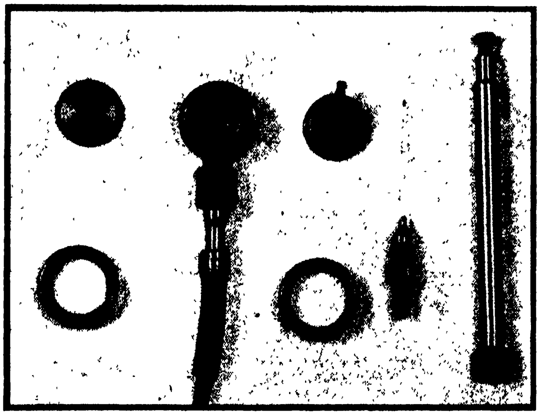


FIG. 19. Pressure gauge with attachments. This is the gauge of Fig. 18 with the tubing connector (left), the liquid system, the needle and its guard (right), removed. The one-half-inch diameter active diaphragm surface is seen as a central ring on the face of the pickup unit. All parts and the gauge itself are made of stainless steel.

circuit depends on the application; several satisfactory ones have been published.[8, 9, 12, 13] One requirement for biological work is that the electrical system record sustained capacitance changes, i.e., record the direct current component as well as the alternating current ones. Cathode-ray oscillographs and oscillograph galvanometers have both been used with success for recorders.

---

[12] D. W. Dana, Rev. Sci. Inst. 5, 38 (1934).

[13] C. H. Brookes-Smith, J. Sci. Inst. 16, 361 (1939).

## APPLICATIONS OF THE GAUGE TO MECHANICAL MEASUREMENTS IN BIOLOGICAL SYSTEMS

Various applications of the gauge to pressure recording during the war are presented in Figs. 21 through 25. Since the applications are described under the figures, no further comments will be made here. Other uses of the gauge as an ultramicrometer or a dilatometer in biological systems might be mentioned. The gauge may be used to measure the movements of the body resulting from recoil caused by heart and blood motions in an electrical ballistocardiograph. An obvious use is as a high speed, isometric "lever" for recording the tension developed by muscles during fast or sustained contractions; it may be possible to use it for single muscle fibers. As a dilatometer, the gauge can be used in plesthysmography, thermometry, and respirometry. The volume changes in muscle during fast contractions may possibly be recorded. It may be used in the measurement

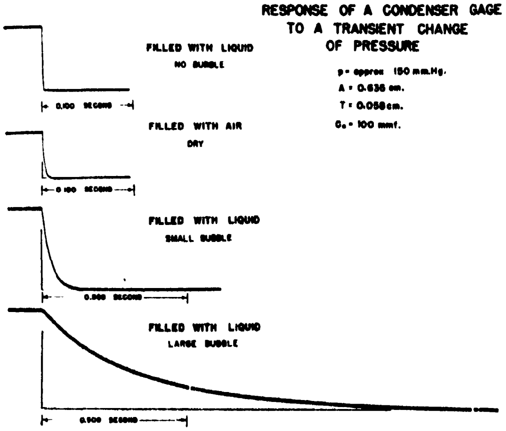


FIG. 20. Response of a capacitor gauge with attached needle to a sudden fall in pressure level. The gauge itself is of the type shown in Figs. 18 and 19. The attached hypodermic needle was a number 23, one-inch long. The volume of the liquid system is 0.4 cc. The needle was immersed in liquid or in air in a short ½-inch diameter well in the face of a 6-inch diameter, horizontal plate, over which was stretched a thin paper diaphragm. The pressure in the air space between the plate and the paper diaphragm was raised with an air supply, and the paper diaphragm was "exploded" away with a mouse trap lever traveling at high speed. The paper broke completely away over the whole plate surface, causing a rapid fall in pressure over the well containing the needle. The conditions causing the various rates of fall of the gauge response are given in the figure. The top record shows a response time of about one-thousandth of a second, which is an upper limit, presumably set by a time constant in the electrical circuit used for recording, rather than by the response time of the mechanical system. (To fill the system with liquid without bubbles, the contained air was flushed out with pure carbon dioxide gas before filling; any bubbles formed of $CO_2$ quickly dissolve in the liquid, leaving the system completely liquid filled.)

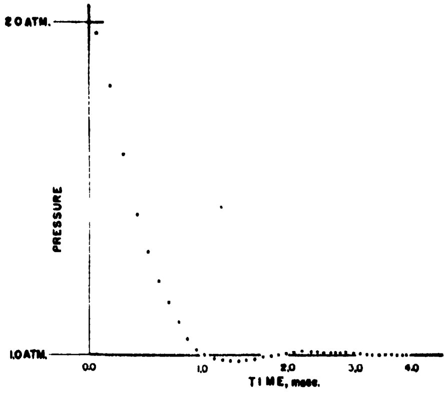


FIG. 21. Explosive decompression of a short cavity. The condenser pressure gauge (of the type shown in Fig. 17) was placed in the bottom of a well 3 inches long and ¾-inch in diameter and containing air. By the technique described under Fig. 20, the pressure at the mouth of this well was suddenly dropped to the ambient pressure from twice the ambient value. This retouched record of the pressure was photographed on a cathode-ray oscilloscope. The time between dots is one ten-thousandth of a second. Note the relatively small after-vibrations in contrast with those of Fig. 22. This gauge was subsequently used in recording the pressure fall in explosively decompressed pressurized cabins in an altitude chamber; the minimum time of pressure fall recorded in these cabins was about four times the value shown here.

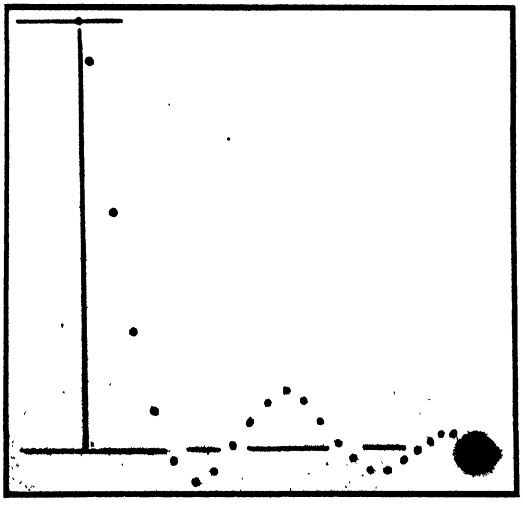


FIG. 22. Explosive decompression of a long air cavity. The technique and the pressure range are the same as those of Fig. 21. The ¾ inch diameter cavity in this case, however, is 3 feet long. The time between dots is two ten-thousandths of a second. The amplitude of the first after-vibration is 2.0 pounds per square inch, after an initial fall of 15 pounds per square inch. This record shows some of the artifacts created by trying to measure relatively large rapid changes in pressure through a long pressure lead, filled with air.

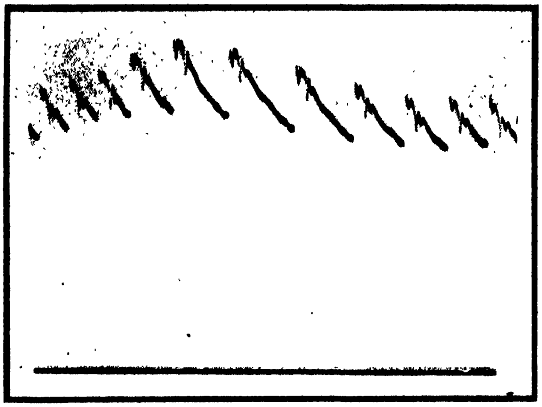


FIG. 23. A blood pressure record in the carotid artery of a dog. The pressure base line is at the bottom of the figure. The peak pressure is approximately 150 mm Hg. The record was taken with a pressure gauge of the type shown in Figs. 18 and 19. This record is shown through the kindness of H. C. Bazett.

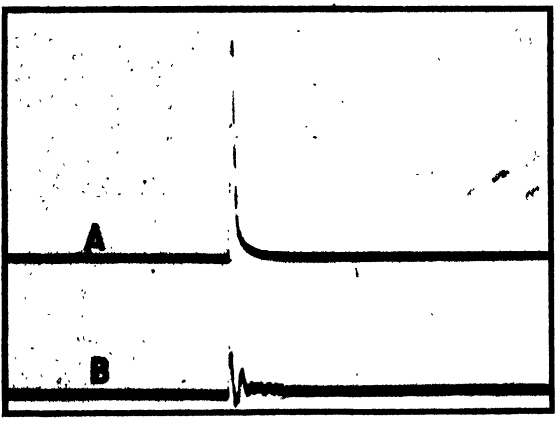


FIG. 24. Record of the intrathoracic and the intraspinal pressure in an explosively decompressed dog. $A$ is the intrathoracic pressure record, and $B$ the intraspinal pressure record. The breaks in the falling phase of record $A$ are at 1/120 second intervals. The condenser gauge used was of the type shown in Figs. 18 and 19. The dog was in a small chamber at an equivalent pressure of 8000 feet; a thin film diaphragm separated this chamber from a larger one at a lower pressure. When the diaphragm was ruptured, the air pressure around the dog fell in 0.03 seconds to that equivalent to 50,000 feet. The two gauge pick-up units were liquid filled, with an electrode side open to ambient pressure and its changes (Fig. 16). The needles were in their respective body spaces. The pressure rise with respect to ambient (about 30 mm Hg in $A$) is apparently due to the relative delay of the air flowing from the lungs. This figure is reproduced through the courtesy of Fred A. Hitchcock, Ohio State University.

of the stimulus of nerve or muscle when pressure or mechanical stresses are applied in a slow or

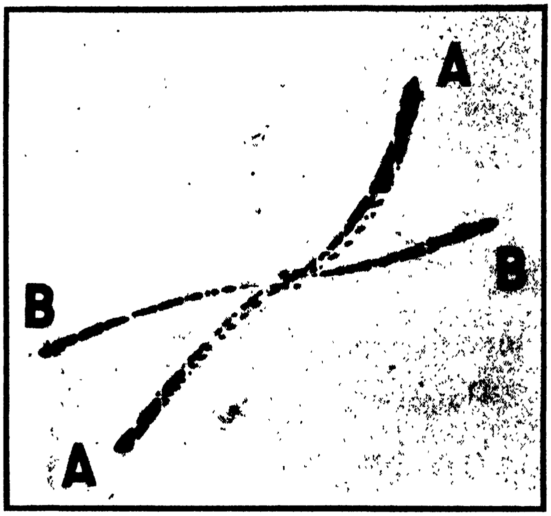


FIG. 25. Pressure-flow curves for air through the nasal passageway. Two, flexible membrane-type pressure gauges with a pressure range of 100 mm of water and 10 mm of water for a capacitance change of 5 mmf were used for recording air pressure and air flow, respectively. The record is a photograph of a cathode-ray tube screen, with flow appearing on the horizontal axis and pressure on the vertical axis. Zero flow and zero pressure are at the center of the record where the two lines cross. Expired gas flow and pressure are in the right upper quadrant, inspired values in the lower left one. The flow values cover about ±50 liters/minute; the pressure about ±80 mm of water. The flow is recorded by measuring the pressure drop across a laminar-flow type glass wool resistor, placed in the tube exit from an oxygen mask covering the subject's nose and mouth. The pressure was measured differentially between the inside of the mask and in $A$–$A$, the subject's oropharynx, and in $B$–$B$, his nasopharynx. In both cases, one nostril is plugged; in $A$–$A$ the pressure tap tube goes into the mouth; in $B$–$B$ the tube goes through the plugged nostril. Record courtesy of Forman and Benevides.

in a fast manner. It is hoped that the design theory presented above will be an aid in extending the applications of this type of gauge in biological research and in the engineering and physical fields.

## ACKNOWLEDGMENTS

The authors wish to express their thanks to Dr. D. W. Bronk for his encouragement during the course of this work; to A. J. Rawson and J. P. Hervey for their help and suggestions; and to J. R. Pappenheimer and G. A. Millikan for their stimulating applications for this gauge.

# Conductivity of Metallic Surfaces at Microwave Frequencies[*]

E. MAXWELL
*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Received February 4, 1947)

Methods of measuring effective conductivities at microwave frequencies are described. These consist of either measuring the transmission loss in a long waveguide, or in measuring the $Q$'s of resonant cavities. Both methods have been applied to measurements at 1.25 cm. Results for a number of metals are presented. Deviations from d.c. conductivity are thought to be due to surface roughness.

## 1. INTRODUCTION

IN the course of development of microwave components it was early observed that the high frequency conductivity of many metallic surfaces was frequently lower than the d.c. conductivity. This phenomenon was especially noticeable in the 1.25-cm region. The work described here was done for the purpose of collecting data on the conductivities of representative surfaces in the region of 1.25 cm.

The r-f conductivity is important in at least two applications. These are:

(1) Long transmission systems where the total attenuation is governed by the conductivity.

(2) Cavity resonators, in which the maximum attainable $Q$ is limited by the effective conductivity.

The effective conductivity of metals at microwave frequencies is governed by a superficially thin surface layer. The current density inside the metal decreases exponentially with depth according to $J=J_0\exp-[(\omega\mu\sigma/2)]^{\frac{1}{2}}x$, where $J_0$ = current density at the surface, $\omega$ = angular frequency, $\mu$ = permeability, $\sigma$ = conductivity, and $x$ = distance into conductor measured along the normal.

When $x=(2/\omega\mu\sigma)^{\frac{1}{2}}$, $J=J_0/e$. This value of $x$ is defined as the skin depth $\delta$. It is easy to show that the total loss, obtained by integrating $J^2/2\sigma$ with respect to $x$, (between 0 and $\infty$), is equal to $J_0^2/2\sigma\delta$. For purposes of calculating losses one may consider the current confined to a skin of thickness $\delta$ within which the current density is uniform and equal to $J_0$. Actually of course 37 percent of the total current flows at depths below $\delta$.

This formulation is rigorously correct only for an infinite plane surface. However, it remains an excellent approximation for curved surfaces whose radii of curvature are large compared to $\delta$. In computing losses one evaluates the integral

$$\int\int_S \frac{J_0^2}{2\sigma}ds=\int\int_S \frac{H_t^2}{2\sigma}ds$$

taken over the surface of the waveguide or cavity. $H_t$ is the tangential component of magnetic field. (The factor 2 comes in since we are interested in average power dissipation, and $J_0$. and $H_t$ are both maximum values of sinusoidal time functions.) If the metal surface is not smooth but has surface irregularities and scratches whose dimensions are comparable with $\delta$, the foregoing treatment is evidently an over-simplification. The presence of scratches may alter the current distribution especially if the scratches are in a direction orthogonal to the lines of current flow. The scratches could conceivably act like constrictions in the effective cross section, thereby increasing the resistance.

## 2. MEASURING TECHNIQUES

Two general methods of measurement have been employed. In the first method the sample to be measured is in the form of a long piece of waveguide or equivalent transmission system. If one end of the waveguide is shorted, the standing-wave ratio seen looking into the other end is a known function of the guide attenuation, which in turn depends on the conductivity. Thus, from a single measurement of standing-wave

---

[*] The research reported in this paper was made possible in part through support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the Army Signal Corps, the Navy Department (Office of Naval Research), and the Army Air Forces (Air Materiel Command), under the Signal Corps Contract No. W-36-039 sc-32037.

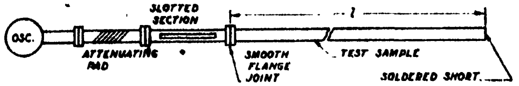


FIG. 1. Attenuation measurement by long guide technique. Rectangular guide.

ratio, the effective conductivity may be computed.

In the second method, the sample forms the walls of a resonant cavity. The $Q$ of the cavity is a known function of the conductivity, and hence the conductivity may be calculated from the measured $Q$. A variation of this method was found quite useful. The input conductance at resonance is directly proportional to the losses in the cavity, or more precisely, to the square root of the wall resistivity. The constant of proportionality depends on the coupling and, if known, a single measurement of the input standing-wave ratio at resonance permits one to calculate the resistivity.

### 2.1 DETAILS OF FIRST METHOD

The measurement set-up for rectangular wave guide is illustrated in Fig. 1. If $r$ is the measured voltage standing-wave ratio, and $l$ is the length of the wave guide, the attenuation is given as follows:

$$\text{Attenuation per unit length} = \alpha$$

$$= \frac{10}{l}\log_{10}\left(\frac{r-1}{r+1}\right)\ \text{db}$$

$$= \frac{1}{2l}\log_e\left(\frac{r-1}{r+1}\right)\ \text{nepers.}$$

The attenuation and conductivity are related by the following expression:

$$\alpha = \frac{1}{2b}\cdot\frac{1}{[1-(\lambda_0/\lambda_c)^2]^{\frac{1}{2}}}\cdot\left(\frac{4\pi}{\lambda_0\mu c\sigma}\right)^{\frac{1}{2}}\cdot\left[1+\frac{2b}{a}\left(\frac{\lambda_0}{\lambda_c}\right)^2\right],$$

where $\alpha$ = attenuation in nepers per meter, $\sigma$ = conductivity in mhos per meter, $a$ = wave guide width in meters, $b$ = waveguide height in meters, $\lambda_0$ = free space wave-length in meters, $\lambda_c = 2a$ = guide cut-off wave-length in meters, $\mu_0$ = permeability of free space (M.K.S.) $= 4\pi \times 10^{-7}$, and $c$ = velocity of light $= 3\times 10^8$ meters per second. This expression is for the $TE_{10}$ mode only. For $0.170'' \times 0.420''$ wave guide at 1.25 cm, the formula reduces to:

$$\alpha = 299/(\sigma)^{\frac{1}{2}}\ \text{nepers per meter.}$$

The length of waveguide is conveniently chosen so as to have a voltage standing-wave ratio of about 5 or 6. At 1.25 cm with common materials, this usually means a waveguide length of the order of six feet. The correction for losses in the slotted section is then negligible.

The far end of the waveguide should be shorted by means of a soldered plate. A choke plunger is unsatisfactory for this purpose, since some power may leak past. If the waveguide is suspected of inhomogeneities or bad spots, a quarter wave-length should be removed from the shorted end and the test repeated. This will shift the standing-wave pattern by a quarter wave-length, and if bad spots are present, the measured loss will probably be different in the two cases. The loss in the short circuiting plate is of course negligible in comparison to that in the wave guide itself.

The foregoing procedure is straightforward enough when dealing with rectangular wave guide. Frequently, however, one wishes to use circular waveguide. This can be done by using a round-to-rectangular transition piece, or transformer, between the round pipe and the slotted section, but a special technique must be employed to avoid errors due to elliptical polarization in the pipe.

It is inevitable that any long piece of round pipe will have both ellipticity and skewness which will cause the incident wave to split up into a pair of cross-polarized waves. Each of these cross-polarized waves will be reflected at the short and come back down the pipe, but only the vertical component of each will couple into the rectangular waveguide. This will cause the observed standing-wave ratio in the rectangular slotted section to be less than the true standing-wave ratio at the sending end of the round pipe.

This difficulty may be eliminated as follows. It may be shown theoretically that given any skewed elliptical pipe, it is possible to choose a direction of polarization for the incident wave, such that the reflected wave arrives with that same polarization. The reflected wave then couples completely into the rectangular wave guide. and the standing-wave ratio observed in the slotted section is the true standing-wave ratio.

If the pipe is a perfect elliptic cylinder, it is clear that either of the principal axes of the ellipse may be taken as the preferred direction. An actual pipe is more apt to resemble a twisted cylinder in which both the orientation of the axes and the ellipticity are functions of position along the axis of propagation. Thus there may be cross-coupling between the modes, and the reflected wave will not generally have the same orientation as the incident wave.

In practice the round wave guide is arranged so that it may be rotated with respect to the transformer (see Fig. 2). The wave guide is rotated until the position of the maximum standing-wave ratio is found. This is the condition for which the reflected wave returns with the same polarization as the incident wave. A convenient type of round-to-rectangular transformer is a quarter wave-length of oval cross section intermediate in shape between the round and rectangular sections. The loss in such a transformer is quite negligible.

This technique was applied in measuring the effective conductivity of mercury. A circular waveguide was constructed by immersing a polystyrene rod in mercury. The details of this experiment are described in the Appendix.

### 2.2 DETAILS OF SECOND METHOD

In any resonant cavity, the unloaded $Q$ is given by

$$Q_0 = \text{const.} \times \frac{\lambda}{\delta},$$

where $\lambda$ is the resonant wave-length, and $\delta$ is the skin depth. The constant of proportionality is a function of the resonator geometry and is known for many of the simpler cases. Thus if one measures $Q_0$, it is simple to compute the effective conductivity from $\delta$ (if the resonator geometry is one for which the theoretical problem has been solved).

In practice, one couples a waveguide into the cavity by means of a coupling window of some sort and measures the variation of input standing-wave ratio with frequency. The half width of the resulting resonance curve determines $Q_L$, the loaded $Q$. Thus,

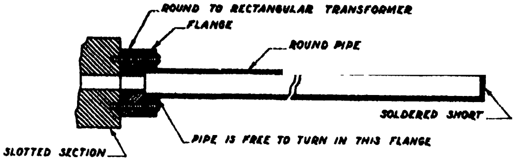

ROUND TO RECTANGULAR TRANSFORMER FLANGE
ROUND PIPE
SOLDERED SHORT
PIPE IS FREE TO TURN IN THIS FLANGE
SLOTTED SECTION

FIG. 2. Attenuation measurement by long guide technique. Round guide.

$$\frac{1}{2Q_L} = \frac{\omega_{\frac{1}{2}} - \omega_0}{\omega_0} \tag{1}$$

where $\omega_0$ is the resonant frequency and $\omega_{\frac{1}{2}}$ the frequency corresponding to the half-power point. (The half-power point is the frequency at which the power absorbed by the cavity is half the power absorbed at resonance.) The ordinates of the half-power points are given by

$$\beta_{\frac{1}{2}} = \frac{\beta_0 + 1 + (\beta_0^2+1)^{\frac{1}{2}}}{\beta_0 + 1 - (\beta_0^2+1)^{\frac{1}{2}}}, \tag{2}$$

where $\beta_0$ = the voltage standing wave ratio at resonance, and $\beta_{\frac{1}{2}}$ = the voltage standing wave ratio at the half-power point.

The loaded and unloaded $Q$'s are related by

$$\frac{Q_0}{Q_L} = 1 + \frac{Y_0}{G}, \tag{3}$$

where $Y_0$ = the line admittance, and $G$ = the input conductance of the cavity. If the cavity is undercoupled,

$$Y_0/G = 1/\beta_0 \tag{4a}$$

and if overcoupled

$$Y_0/G = \beta_0. \tag{4b}$$

The existence of undercoupling or overcoupling is determined experimentally by the presence of a minimum or a maximum, respectively, at the input window.

A calculating technique which makes somewhat better use of the data is as follows: The equation of the resonance curve is:

$$\left(\frac{2\Delta Q_0 G}{Y_0}\right)^2 \frac{Y_0}{G} + \frac{Y_0}{G} + \frac{G}{Y_0} = \frac{1+\beta^2}{\beta}, \tag{5}$$

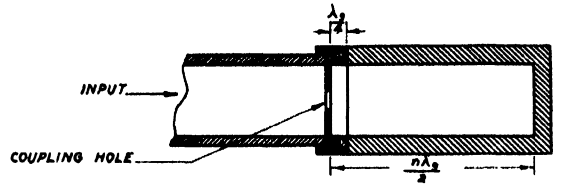


FIG. 3a. Two part resonator.

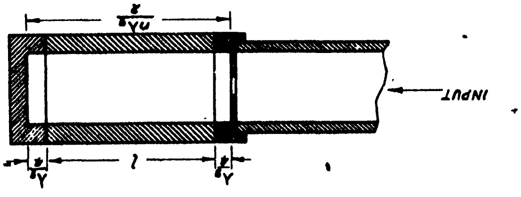


FIG. 3b. Three part resonator.

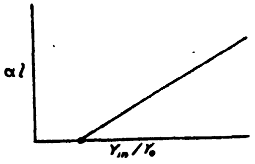


FIG. 3c. Calibration plot.

where $\Delta=(\omega-\omega_0/\omega_0)$, and $\beta$ = voltage standing wave ratio at frequency $\omega$. If one plots $(1+\beta^2/\beta)$ *vs.* $\Delta^2$, a straight line is obtained. By fitting the best straight line to the experimental data, one may evaluate $Q_0$, $Q_L$, and $Y_0/G$. This procedure makes use of the entire resonance curve and not merely three points.

It is thus clear that one may determine $Q_0$ from bandwidth measurements and, having found it, calculate the corresponding effective conductivity of the wall material. In principle then, one may construct a series of cavities from the various materials to be tested and evaluate the effective conductivities by the methods described above.

Although this technique is straightforward, there were a number of practical reasons which militated against its use in this investigation. First of all, the active surfaces, being the interior walls of the cavity, would not always be readily accessible for polishing, plating, or other surface treatment. Secondly, the bandwidth measurements were tedious and difficult with the techniques then available, and since many measurements were to be made, they would have proved very time consuming. The bandwidth measurements require stable oscillators and means for measuring accurately small frequency intervals. Although such equipment is available today, it was not at the time this work was begun. Therefore, a modified technique was developed in which a minimum of bandwidth measurements was required.

Suppose we consider a cavity which is feeding power back to the line, instead of absorbing power from the line. This would be the case when the oscillator is shut off, and the field in the cavity starts to decay. Then,

$$Q_0=2\pi\frac{\text{Energy stored in the cavity}}{\text{Energy dissipated in cavity walls per cycle}},\tag{6a}$$

$$Q_L=2\pi\frac{\text{Energy stored in the cavity}}{\text{Energy dissipated in cavity walls}+\text{energy radiated back to line per cycle}},\tag{6b}$$

$$Q_W=2\pi\frac{\text{Energy stored in the cavity}}{\text{Energy radiated back to line per cycle}}.\tag{6c}$$

$Q_W$ is the window $Q$, or external $Q$. It depends only on the shape of the cavity and the geometry of the coupling window and not upon the conductivity of the walls.

Evidently,

$$1/Q_L=(1/Q_0)+(1/Q_W).\tag{7}$$

Also since

$$Q_0/Q_L=1+(Y_0/G),\quad Y_0/G=Q_0/Q_W.\tag{8}$$

Equation (8) means that the standing-wave ratio at resonance, or its reciprocal, is proportional to the unloaded $Q$.

Thus in principle, if we have a series of cavities which are identical in all dimensions, including those of the coupling window, the input standing-wave ratio of each cavity will be proportional to its $Q$, and the constant of proportionality is the same for all cavities. A bandwidth measurement on just one of the cavities will suffice to determine the window $Q$, and thus the constant of proportionality. The fact is, however, that the window $Q$ varies very rapidly with the coupling window dimensions, so that it is impracticable to maintain the necessary tolerances.

A more practical approach is to build a cavity in two or more demountable sections, such that the test piece may form one of these sections. Examples of this type of construction are indicated in Figs. 3–6.

The first one tried was the $TE_{011}$ resonator of Fig. 5a. The removable sample is the end plate. The lines of current flow are coaxial circles and nowhere cross the contact surface, so that the losses in the oscillating mode are not affected by the contact resistance. There is considerable advantage in having the sample in the form of a flat plate, since this is the most convenient form for cleaning, polishing, plating, etc. This resonator had two serious disadvantages, however. First of all, the $TE_{011}$ mode is intrinsically degenerate with the $TM_{111}$ odd and even modes, so that extreme care must be taken to avoid exciting these modes (if it can be avoided at all). Secondly, there are a number of non-resonant modes present in the cavity which influence the window $Q$ in an undesirable fashion. This trouble seems to arise whenever we use a resonator whose dimensions are such that more than one mode can propagate in the cavity, even though the unwanted modes are non-resonant. For the cavity shown in Fig. 5a, the $TE_{11}$, $TM_{01}$, and $TE_{21}$ modes, which are lower than the $TE_{01}$, will certainly propagate along the axis of the cylinder, and in addition many higher ones will also propagate because of the relatively large diameter. (A squat cavity was chosen in order to make the ratio of sample loss to total loss as large as possible.) These extraneous modes are excited because the presence of the coupling hole, or slit, imposes boundary conditions which cannot be satisfied by the resonant mode alone. The currents which flow, as a result of these modes, do cross the contact surfaces and are therefore controlled by the contact resistance.

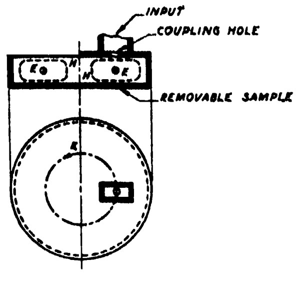


FIG. 5a. $TE_{011}$ mode resonator.

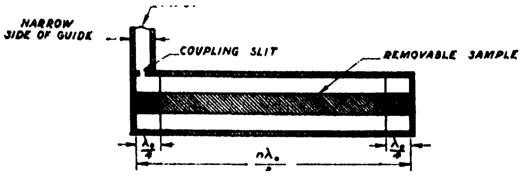


FIG. 5b. Coaxial resonator—TEM mode.

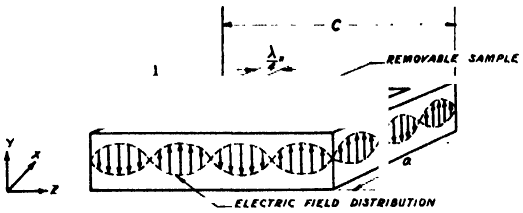
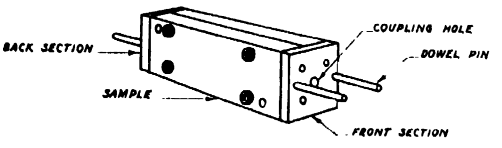
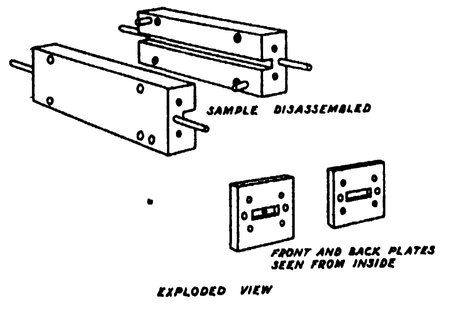


FIG. 6. Rectangular resonator.



PICTORIAL VIEW



EXPLODED VIEW

FIG. 4. Rectangular resonator details.

The energy stored in these modes determines the equivalent susceptance of the coupling iris and hence the window $Q$. In other words, the window $Q$ depends upon the contact resistance which is not a reproducible quantity. Therefore, the window $Q$ varies from measurement to measurement and must be redetermined each time. Since it was desired to eliminate bandwidth measurements, this type of resonator was discarded.

There is, however, considerable merit in using flat plate samples. With the stabilized oscillators now available,[1] bandwidth measurements in the region of 24,000 Mc are no longer difficult so that the method of the preceding paragraph is entirely practicable today. The $TE_{011}$ cavity is not the best choice, however, owing to the degeneracy with the $TM_{111}$ modes. A suggested cavity, using a removable flat plate sample, is illustrated in Fig. 6. This would operate in the $TE_{l0n}$ mode. By choosing the dimensions such that $b<\lambda_0/2$, and if $a$ and $b$ are incommensurable, one may avoid accidental degeneracies. The resonant wave-length $\lambda_0$ is determined from the relation $(2/\lambda_0)^2=(l/a)^2+(n/c)^2$ in which $l$ and $n$ are equal to the number of half-period variations along the $x$ and $z$ axes, respectively. The contact surfaces are at current nodes. A cylindrical cavity in the $TM_{0m0}$ mode could be used as well, in which case the sample plate would be circular.

Figure 5b illustrates another type of demountable cavity in which the sample is the removable center conductor. This resonator operates in the principal mode. The breaks in the inner and outer conductors occur at points of zero current. One version of this was tried in which the diameter was large enough so that the $TE_{11}$ mode could propagate. Here again it was found that the window $Q$ was not reproducible.

The cavity finally chosen for making most of the measurements was the demountable rectangular type illustrated in Figs. 3 and 4. This was taken as the best compromise between the conflicting requirements of convenient sample shape and simplicity of measuring technique. The cavity operates in the lowest mode, the $TE_{10n}$. The cross-sectional dimensions are standard waveguide dimensions for 1.25 cm, and therefore modes higher than the $TE_{10}$ cannot propagate. The higher modes generated in the vicinity of the coupling window are damped out very rapidly. The breaks in the walls occur across lines of no current flow and are sufficiently removed from the neighborhood of the coupling window so that the higher mode amplitude is practically nil.

Both two- and three-section cavities were used, although the three-section type is preferred because it is easier to get at the active surfaces. An exploded view of the three-part resonator is shown in Fig. 4.

The normalized input conductance at resonance is:

$$G/Y_0=Q_w/Q_0. \tag{8a}$$

For the simple rectangular resonators used this may be written in the form

$$\frac{G}{Y_0}=\frac{n\pi B^2}{2Q_0}\left(\frac{\lambda_g}{\lambda_0}\right)^2=\left(\frac{B}{Y_0}\right)^2(\alpha l+R), \tag{9}$$

where $n=$ the axial length in numbers of half wave-lengths, $\lambda_g=$ the guide wave-length, $\lambda_0=$ the free space wave-length, $B=$ the susceptance of the coupling window, $\alpha=$ the attenuation in the sample portion in nepers/meter, $l=$ the length of the sample portion, and $R=$ a resistance which represents the dissipation in the front and back quarter-wave sections.

In the above formula $(\alpha l)$ is the total attenuation in the sample. The relation between $(\alpha l)$ and $G/Y_0$ is linear. A convenient way to calibrate a pair of front and back quarter-wave sections is as follows. The attenuation in a long piece of waveguide is measured by the short-circuited guide technique described earlier in this paper. Samples with different values of $(\alpha l)$ are obtained by cutting this piece up into short sections, each one of which is a different multiple of a half wave-length. From the data taken with these samples a calibration plot of total attenuation in the sample *vs.* input conductance is obtained as in Fig. 3c.

This calibration plot is then used to determine the attenuation in unknown samples. The sample

---

[1] R. V. Pound, Radiation Laboratory Report 662; R. V. Pound, Radiation Laboratory Report 837; and R. V. Pound, "Electronic Frequency Stabilization of Microwave Oscillators," Rev. Sci. Inst. 17, 490 (1946).

TABLE I. Conductor losses in standard K-band waveguide.[1]

| Material | Measured atten. in db/meter | Calculated[2] atten. in db/meter | Meas. atten. / Calc. atten. | Eff. cond. at K-band in $10^7$ mho/m | d.-c. Cond. in $10^7$ mho/m |
|---|---|---|---|---|---|
| Aluminum[7] | | | | | |
| Pure, commercial (machined surface) | 0.58 | 0.455 | 1.27 | 1.97 | 3.25 (measured) |
| 17*S* Alloy[7] (machined surface) | 0.75 | 0.586 | 1.28 | 1.19 | 1.95 (measured) |
| 24*S* Alloy (machined surface) | 0.66 | 0.635 | 1.04 | 1.54 | 1.66 (measured) |
| Brass | | | | | |
| Yellow (80–20) drawn wave guide | 0.68 | 0.653 | 1.04 | 1.45 | 1.57 (measured) |
| Red (85–15) drawn wave guide | 0.55 | | | 2.22 | |
| Yellow round drawn tubing[3] | 0.90 | 0.844 | 1.07 | 1.36 | 1.56 (measured) |
| Yellow (80–20) (machined surface) | 0.76 | 0.653 | 1.15 | 1.17 | 1.57 (measured) |
| Free machining brass (mach. surface) | 0.75 | 0.673 | 1.12 | 1.11 | 1.48 (measured) |
| Cadmium plate | 0.79-0.87 | 0.711 | 1.13-1.22 | 1.04-0.89 | 1.33 Hdbk. of Phys. and Chem. |
| Chromium plate, dull[4] | 0.67-0.82 | 0.418 | 1.60-1.96 | 1.49-0.99 | 3.84 Hdbk. of Phys. and Chem. |
| Copper | | | | | |
| Drawn O.F.C. wave guide | 0.41 | 0.350 | 1.17 | 4.00 | 5.48 (measured) |
| Drawn round tubing | 0.52 | 0.496 | 1.05 | 4.10 | 4.50 (measured) |
| Machined surface[7] | 0.38 | 0.349 | 1.09 | 4.65 | 5.50 (measured) |
| Copper plate | 0.54-0.61 | 0.337 | 1.60-1.81 | 2.28-1.81 | 5.92 Hdbk. of Phys. and Chem. |
| Electroformed wave guide[7] | 0.46 | 0.337 | 1.37 | 3.15 | 5.92 |
| Gold plate | 0.60 | 0.404 | 1.48 | 1.87 | 4.10 Hdbk. of Phys. and Chem. |
| Iron, electroformed[7] | 1.01 | | | | |
| Mercury[6] | 2.50 | 2.54 | 0.98 | 0.104 | 0.104 Hdbk. of Phys. and Chem. |
| Monel (machined surface)[5,7] | 2.08 | 2.07 | 1.01 | 0.155 | 0.156 (measured) |
| Nickel[7] | | | | | |
| Electroformed wave guide | 0.82 | | | | |
| Nickel plate | 1.11 | | | | |
| Silver | | | | | |
| Coin silver drawn wave guide | 0.45 | 0.375 | 1.20 | 3.33 | 4.79 (measured) |
| Coin silver lined wave guide | 0.60 | 0.375 | 1.60 | 1.87 | 4.79 (assumed) |
| Coin silver (machined surface)[7] | 0.51 | 0.375 | 1.34 | 2.66 | |
| Fine silver (machined surface)[7] | 0.48 | 0.330 | 1.45 | 2.92 | 6.14 Hdbk. of Phys. and Chem. |
| Silver plate | 0.41-0.57 | 0.330 | 1.24-1.73 | 3.98-2.05 | |
| Solder, soft[7] | 1.05 | 0.978 | 1.08 | .600 | 0.70 (measured) |
| Steel, cold rolled (mach. surface) | 2.85 | | | | |

[1] Unless otherwise noted, figures are for $TE_{10}$ mode, $\lambda = 1.25$ cm in rectangular guide of dimensions of 0.170″×0.420″.

[2] Theoretical attenuation for $TE_{10}$ mode in 0.170″×0.420″ rectangular waveguide at 1.25 cm is $2590\sigma^{-\frac{1}{2}}$ db/m where $\sigma$ is given in mho/meters. For the $TE_{11}$ mode in round pipe at 1.25 cm, the attenuation is $3330\sigma^{-\frac{1}{2}}$ db/m.

[3] 0.345″ I.D., $TE_{11}$ mode.

[4] No nickel undercoat.

[5] This surface was somewhat rougher than most machined surfaces.

[6] For method of measurement refer to the Appendix. Figures are expressed for a guide of 0.170″×0.420″ cross section.

[7] Only one sample of these was tested.

may be any multiple of a half wave-length. However, for any given window susceptance there is a range of sample attenuations which produce conveniently measurable standing-wave ratios. It is a good idea to plate the front and back sections with some corrosion-resistant material to insure permanence of calibration.

### 3. EXPERIMENTAL RESULTS

The more important available data on conductor losses at 1.25 cm are summarized in Table I. These are expressed in terms of attenuation of the $TE_{10}$ mode in standard 0.170″×0.420″ waveguide at 1.25 cm, or in some cases for the $TE_{11}$ mode in round pipe of 0.345″ diameter. The first column records the value of the experimentally measured attenuation, the second that of the theoretically computed attenuation, and the third column gives the ratio of the two; the fourth and fifth columns list the effective conductivity at 1.25 cm and the d.c. conductivity, respectively.

The theoretical attenuations are computed by means of the formulas given in the footnotes of Table I. The conductivities used in these formulas are the d.c. conductivities listed in the fifth column. The d.c. conductivity of the samples was measured wherever possible but in some cases it was not convenient or possible to do so. In these other cases, the sources from which the d.c. conductivity data were taken, are indicated. In comparing theoretical and experimental performance it is important to have accurate data on the d.c. conductivities if the comparison is to have much significance. In the case of fairly pure materials, such as mercury or electrodeposited metals, the conductivity figures given in handbooks for pure metals are probably good enough, but for commercial metals and alloys there may be discrepancies between tabulated and actual values.

Although the listing is far from complete, many of the common conductor materials are present. There is usually some variation in conductivity among different specimens of the same material, particularly in machined and electroplated surfaces. Where the variation among specimens was more than a few percent, the outside limits of attenuation are stated in the table. Some of the figures are based on a single specimen and are so indicated. In the other cases anywhere from two to about six samples were tested. The probable error to be associated with the attenuation figures is estimated as less than two percent. In the case of the soft solder the error may be of the order of 10 percent since the measurement quoted here is of an earlier vintage than the others and the technique employed was less accurate.

The surfaces examined fall into three categories: drawn, machined, and electrodeposited. The drawn surfaces are as a rule quite good and although their losses are greater than theory predicts, the discrepancy is not large. In the case of brass it may amount to only 5 percent. Machined surfaces are frequently poorer, while plated surfaces seem to vary a good deal.

It seems reasonable to assume that the increase in r-f conductivity over the d.c. value is due primarily to surface roughness. It is difficult, however, to establish such a correlation on the basis of the observed data since no reliable index of surface roughness is available for the samples tested.

The measurement of the r-f conductivity of mercury, which was briefly mentioned earlier, is of interest in this connection. In this experiment the attenuation was measured in a waveguide made by immersing a smooth polystyrene rod in mercury. Because of the surface tension of the mercury, the resulting surface was considered to be free from the usual multitude of scratches and crevices present on solid metallic surfaces. Furthermore, since the skin depth is about eight times as big in mercury as in copper or silver, such irregularities as remain are relatively less important. The measured attenuation agreed with the calculated value within 2 percent which was about as accurately as the standing-wave ratio could be measured. Actually, however, the experimental figure may be in doubt by 4 or 5 percent as a result of a possible uncertainty in the assumed value of the conductivity of coin silver which enters into the experiment. The fact that agreement was attained here between theory and experiment again implies that the lack of agreement observed in other cases is due to surface irregularity. This matter is discussed in the Appendix where the experiment is described in detail.

TABLE II. Conductor losses in K-band waveguide with protective coatings.

| Material | Measured attenuation in db/meter |
|---|---|
| Palladium flash ($10^{-5}$ in.) on coin silver | 0.6 |
| Rhodium flash ($10^{-5}$ in.) on coin silver | 1.0 |
| Sperry f710 lacquer on copper plate | 1.2 |
| Same surface without lacquer | 0.6 |
| Zinc chromate olive drab primer on 17S aluminum alloy | 1.0 |

It should be noted that for the monel sample also, there is substantial agreement between observed and calculated attenuations. The skin depth in monel is approximately the same as in mercury. Thus one may infer that because of the greater skin depth the scratches are less important.

Many attempts were made to improve the conductivity of samples by polishing the active surfaces by both mechanical and electrolytic methods, but none of these were successful. Probably something of the order of a metallographic or optical polish is necessary, but it was not possible to approach this degree of excellency on account of the awkward shape of the surfaces to be polished. In this connection, flat samples would be decidedly advantageous.

The reverse process was possible. The conductivity was definitely lowered in many cases by abrading the surface. For example, when some yellow brass waveguide tubing was broached out, the attenuation was increased from 0.68 db/m to 0.74 db/m.

Thin oxide films are not harmful as long as the oxide resistivity is high. It may be shown that the effect of oxide films is small as long as the skin depth in the oxide material is large compared to the oxide layer thickness.

The effect of a few protective coatings in increasing attenuation is shown in Table II.

Of all the solid metals listed mild steel has the greatest loss. This behavior has been explained by Kittel on the basis of the magnetic properties.[2]

## ACKNOWLEDGMENT

I am indebted to Dr. E. M. Purcell, formerly of the Radiation Laboratory, for many helpful suggestions.

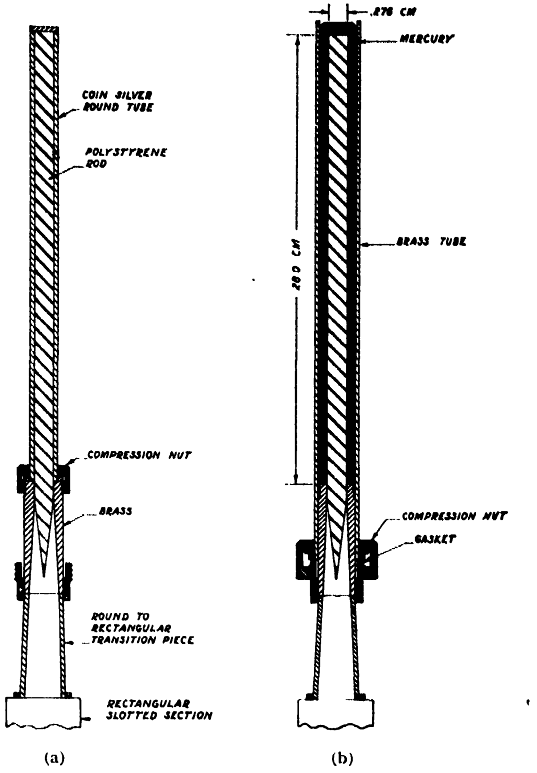
## APPENDIX

### The Mercury Waveguide Experiment

The experimental setup is indicated in Figs. 7a and 7b. The wave guide consisted of a round polystyrene rod immersed in mercury. This was tapered into a round air-filled guide as shown in Fig. 7b. A round-to-rectangular transition coupled the round guide to a standard rectangular slotted section. The rod was then treated as a shorted round waveguide and the attenuation measured by the techniques described under method I for round guide.

The dielectric losses were separated from the conductor losses in the following manner. Before setting up the mercury guide a tightly fitting drawn coin silver tube was slipped over the polystyrene rod, as in Fig. 7a. The input standing-wave ratio was measured, and from this the total attenuation in the polystyrene-silver guide was calculated. The silver tube was then removed and the polystyrene rod immersed in mercury as in Fig. 7b. The standing-wave ratio was again measured and the attenuation in the polystyrene system calculated. The difference between these two figures is the conductor attenuation in the mercury minus the conductor attenuation in the coin silver. The attenuation in the coin silver was calculated by using the effective conductivity established from previous measurements on drawn rectangular coin silver wave guide, and the attenuation due to the mercury was determined. The attenuation thus found agreed with theory within 2 percent.

An estimate of the limit of error was made on the following basis. The attenuation in the polystyrene-silver system was 1.46 db; in the polystyrene-mercury system it was 3.29 db, mercury attenuation less silver attenuation



FIGS. 7a and 7b. Mercury wave guide.

=1.83 db. The effective conductivity of coin silver wave guide is $3.3\times 10^7$ mhos per meter. The conductor attenuation for the $TE_{11}$ mode in the silver tube was 0.40 db for a 28 cm length (see Fig. 7b). The net loss due to the mercury alone was therefore 2.23 db or 8.0 db per meter. The value calculated by using the tabulated figure of $0.1044\times 10^7$ mhos per meter for mercury at 20°C, was 8.10 db per meter. The error depends largely on the accuracy of the assumed attenuation for the coin silver. By assuming a 20 percent error in estimating the attenuation in silver, which should certainly be an outside figure, the uncertainty in the mercury attenuation would be less than 4 percent.

It should also be pointed out that residual mismatch in the taper or transition causes very little error because of the fact that the mismatch appears in both measurements and almost completely cancels out in the final result. As a preliminary part of the experiment, taper and transition mismatch was investigated by slipping

---

[2] C. Kittel, "Theory of the dispersion of magnetic permeability in ferromagnetic materials at microwave frequencies," Phys. Rev. 70, 281 (1946).

some close fitting poly-iron sleeves over the polystyrene rod, in place of the silver tube, so as to obtain a matched absorbing load. The voltage standing-wave ratio looking in was 1.08. It may be verified by calculation that in the worst case this would make an error of less than ½ percent.

*Formula Used in Computing Attenuation:*[3]

$$\alpha=\frac{1}{a}\left(\frac{\pi\epsilon\nu}{\sigma}\right)^{\frac{1}{2}}\left[1-\left(\frac{\nu_{11}'}{\nu}\right)^{2}\right]^{\frac{1}{2}}\times\left[\left(\frac{\nu_{11}'}{\nu}\right)^{2}+\frac{1/u_{11}'^{2}}{1-1/u_{11}'^{2}}\right].$$

$a$ = radius of polystyrene rod, $\epsilon=2.52\epsilon_0$ for polystyrene, $u_{11}'=1.84$, $\nu_{11}'=[u_{11}'/2\pi a(\epsilon\mu)^{\frac{1}{2}}]$, and $\nu$ = frequency. M.K.S. units used.

---

[3] See J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., 1941), p. 544.

---

# Note on Circular Loop Antennas with Non-Uniform Current Distribution

G. Glinski
*Northern Electric Company, Montreal, Quebec, Canada*

(Received February 6, 1947)

In this paper are presented the approximate formulas for the components of radiation vectors of a short-circuited circular loop with non-uniform current distribution. The formulas are valid for the ratio of loop perimeter to wavelength of the order of 0.5 or less, and assume the current distribution of the hyperbolic cosine form. These formulas lead to the radiation intensity formula from which the expressions for the horizontal and vertical field patterns are derived. The latter formulas are further simplified assuming that the attenuation constant is much smaller than the phase constant. From the expression for horizontal field pattern, it follows that the pattern is symmetrical about the loop axis of symmetry. Moreover, this horizontal pattern exhibits a directional effect with maximum field in the direction of the feeder end of the loop. This directional effect is a function of loop dimensions. The theoretical horizontal pattern agrees very closely with the experimental one. From the expression for vertical field pattern, it follows that the non-uniform current distribution produces a pattern intermediate between that for horizontal dipole and horizontal small loop with uniform current distribution. The field intensity in the zenithal direction is again a function of loop dimensions. Using the expression for radiation intensity, formulas for a radius of equivalent circular horizontal field pattern, power gain, average power gain, and radiation resistance are derived. It is shown that the average power gain is essentially a function of loop radius and decreases with the increase of the latter. Finally, the approximate expression for the attenuation constant of the transmission line equivalent to the loop is derived.

The purpose of this note is to present formulas for radiation pattern and gain of circular loop antennas with non-uniform current distribution. Although the problem is of considerable practical importance for the designer of some f-m broadcast antennas[1–4] whose elements are often circular loops or may be considered as such to the first approximation, there appears to be very little in the literature on the subject.[5, 6]

The general problem of current distribution on a circular radiating loop has been treated previously.[7, 8]

In the following, the actual current distribution will be approximated by that on the equivalent uniform transmission line.[9] The effect of radiation will be included in the attenuation of this line.[10]

## GENERAL FORMULA FOR RADIATION VECTOR OF CIRCULAR LOOP

Consider the circular loop in $(\theta=\pi/2)$-plane of the spherical coordinate system $r$, $\varphi$, $\theta$ of

---

[1] N. W. Scheldorf, Gen. Elec. Rev. **46**, 163 (1943).

[2] C. R. Jones, Communications **26**, 36 (1946).

[3] Bell Lab. Record **24**, 163 (1946).

[4] R. F. Holz, F.M. and Television **6**, 45 (1946).

[5] E. M. Williams, Proc. I.R.E. **28**, 480 (1940).

[6] J. B. Sherman, Proc. I.R.E. **32**, 534 (1944).

[7] E. Hallen, Nova Acta Regia Soc. Sci. Upsaliensis, Series IV, **11**, 33 (1938).

[8] F. B. Pidduck, *Currents in Aerials and High-Frequency Networks* (Clarenden Press, Oxford, 1946), p. 62.

[9] L. L. Libby, Proc. I.R.E. **34**, 641 (1946).

[10] S. A. Schelkunoff, *Electromagnetic Waves* (D. Van Nostrand Company, Inc., New York, 1943), p. 200.

Fig. 1. The center of the loop is at the origin of the coordinate system.

Consider two diametrically opposite loop elements $A$ and $B$ with the coordinates $\bar{r}$, $\bar{\varphi}$, $\pi/2$, and $\bar{r}$, $\bar{\varphi}+\pi$, $\pi/2$, and currents of complex amplitudes $I_A$ and $I_B$, respectively.

Following Schelkunoff[11] the components of elementary radiation vectors are

$$\begin{aligned} dN_{A,x} &= -I_A\bar{r}\sin\bar{\varphi}d\bar{\varphi}e^{jt},\\ dN_{A,y} &= I_A\bar{r}\cos\bar{\varphi}d\bar{\varphi}e^{jt},\\ dN_{B,x} &= I_B\bar{r}\sin\bar{\varphi}d\bar{\varphi}e^{-jt},\\ dN_{B,y} &= -I_B\bar{r}\cos\bar{\varphi}d\bar{\varphi}e^{-jt}, \end{aligned} \tag{1}$$

where $t=\beta\bar{r}\sin\theta\cos(\varphi-\bar{\varphi})$.

The components of equivalent elementary radiation vector are therefore

$$\begin{aligned} dN_x &= -I_1\bar{r}\sin\bar{\varphi}d\bar{\varphi},\\ dN_y &= I_1\bar{r}\cos\bar{\varphi}d\bar{\varphi}, \end{aligned} \tag{2}$$

where $I_1=I_Ae^{jt}-I_Be^{-jt}$.

It is convenient to introduce the symmetrical and anti-symmetrical components of currents defined by

$$\begin{aligned} I_A &= (I_s+I_a)/2,\\ I_B &= (I_s-I_a)/2. \end{aligned} \tag{3}$$

In terms of $I_s$ and $I_a$, formulas (2) can be written as:

$$\begin{aligned} dN_x &= -I_1\bar{r}\sin\bar{\varphi}d\bar{\varphi},\\ dN_y &= I_1\bar{r}\cos\bar{\varphi}d\bar{\varphi}, \end{aligned} \tag{4}$$

where $I_1=jI_s\sin t+I_a\cos t$.

In the case of f-m broadcast antennas, $\beta\bar{r}$ is of the order of 0.5. It is quite accurate therefore to replace the trigonometric functions in formula (4) by first terms of their Bessel expansion[12] (Appendix 1). To the first approximation therefore

$$\begin{aligned} dN_x &\cong -I_2\bar{r}\sin\bar{\varphi}d\bar{\varphi},\\ dN_y &\cong I_2\bar{r}\cos\bar{\varphi}d\bar{\varphi}, \end{aligned} \tag{5}$$

where $I_2=J_0(\beta\bar{r}/2)[j4I_sJ_1(\beta\bar{r}/2)\sin\theta\cos(\varphi-\bar{\varphi})+I_aJ_0(\beta\bar{r}/2)]$.

The components of total radiation vector are found by summing up the contributions of all elementary components of radiation vectors:

$$N_x=\int_0^{\pi} dN_x,\quad N_y=\int_0^{\pi} dN_y. \tag{6}$$

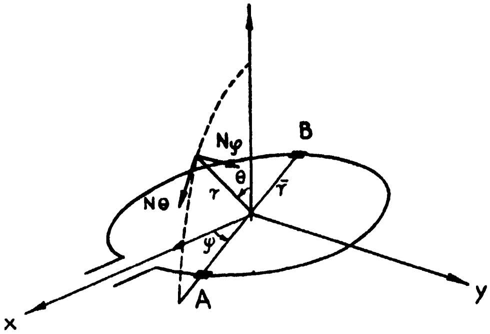


FIG. 1. Coordinate system used.

## RADIATION VECTOR OF SHORT-CIRCUITED CIRCULAR LOOP

Consider the important practical case of the loop fed at $\bar{\varphi}=0$, and short-circuited at $\bar{\varphi}=\pi$. The current distribution on the equivalent transmission line is

$$I_l=I\cosh\gamma l; \tag{7}$$

where $I$ = current at the short-circuit; $\gamma$ = propagation constant $=\alpha+j\beta$; $l$ = distance from the short-circuit. $I_A$ and $I_B$ of formulas (3) are:

$$\begin{aligned} I_A &= I\cosh\gamma\bar{r}(\pi-\bar{\varphi}),\\ I_B &= I\cosh\gamma\bar{r}\bar{\varphi}. \end{aligned} \tag{8}$$

Therefore,

$$\begin{aligned} I_s &= I[(\cosh\gamma\bar{r}\pi+1)\cosh\gamma\bar{r}\bar{\varphi}\\ &\qquad -\sinh\gamma\bar{r}\pi\sinh\gamma\bar{r}\bar{\varphi}],\\ I_a &= I[(\cosh\gamma\bar{r}\pi-1)\cosh\gamma\bar{r}\bar{\varphi}\\ &\qquad -\sinh\gamma\bar{r}\pi\sinh\gamma\bar{r}\bar{\varphi}]. \end{aligned} \tag{9}$$

The elementary integration and reduction (Appendix 2) of formulas (6), using the values of $I_s$ and $I_a$ from formula (9), gives

$$N_x=I\bar{r}J_0\left[-jJ_1\sin\theta\frac{16\sin j\gamma\bar{r}\pi}{j\gamma\bar{r}(4+\gamma^2\bar{r}^2)}\sin\varphi\right], \tag{10}$$

$$\begin{aligned} N_y=I\bar{r}J_0\Bigg[&jJ_1\sin\theta\frac{8(2+\gamma^2\bar{r}^2)\sin j\gamma\bar{r}\pi}{j\gamma\bar{r}(4+\gamma^2\bar{r}^2)}\cos\varphi\\ &-J_0\frac{2j\gamma\bar{r}\sin j\gamma\bar{r}\pi}{1+\gamma^2\bar{r}^2}\Bigg], \end{aligned} \tag{11}$$

where $J_0\equiv J_0(\beta\bar{r}/2)$, $J_1\equiv J_1(\beta\bar{r}/2)$.

The spherical components of total radiation vector are, therefore,[13]

---

[11] S. A. Schelkunoff, reference 10, p. 336.

[12] S. A. Schelkunoff, reference 10, p. 55.

[13] S. A. Schelkunoff, reference 10, p. 234.

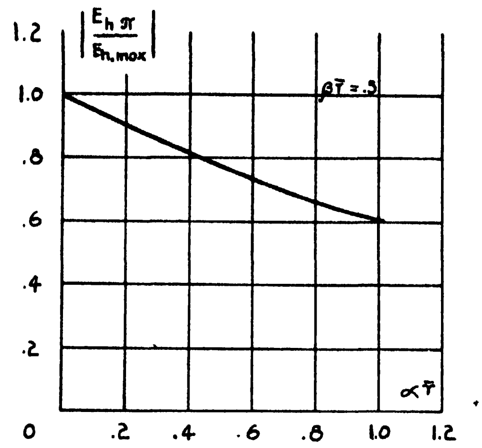


FIG. 2. Ratio of electrical field intensities in ($\varphi=\pi$)- and ($\varphi=0$)-directions in the plane of the loop *versus* attenuation constant $\alpha\bar{r}$ for $\beta\bar{r}=0.5$.

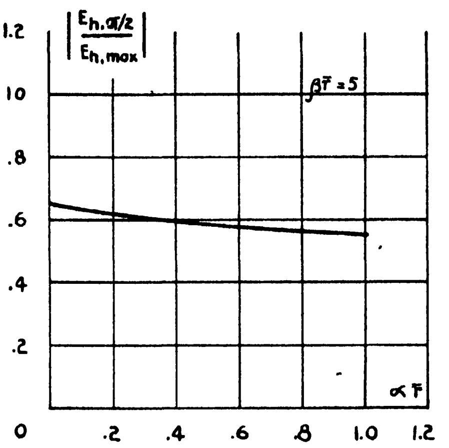


FIG. 3. Ratio of electrical field intensities in ($\varphi=\pi/2$)- (or $3\pi/2$) and ($\varphi=0$)-directions in the plane of the loop *versus* attenuation constant $\alpha\bar{r}$ for $\beta r=0.5$.

$$N_\theta=(N_x\cos\varphi+N_y\sin\varphi)\cos\theta$$

$$=A\left[-j\frac{J_1}{J_0}\frac{\sin 2\theta\sin 2\varphi}{4+\gamma^2\bar{r}^2}-\frac{\cos\theta\sin\varphi}{1+\gamma^2\bar{r}^2}\right];\quad(12)$$

where $A=2I\bar{r}J_0{}^2j\gamma\bar{r}\sin j\gamma\bar{r}\pi$,

$$N_\varphi=-N_x\sin\varphi+N_y\cos\varphi$$

$$=A\left[-j\frac{J_1}{J_0}\frac{4(2+\gamma^2\bar{r}^2\cos^2\varphi)\sin\theta}{\gamma^2\bar{r}^2(4+\gamma^2\bar{r}^2)}-\frac{\cos\varphi}{1+\gamma^2\bar{r}^2}\right].\quad(13)$$

Using the approximations of Appendix 3, the formulas (12) and (13) reduce to:

$$\begin{aligned}N_\theta=A[(&-f(J_1/J_0)\sin 2\theta\sin 2\varphi-a\cos\theta\sin\varphi)\\&+j(-e(J_1/J_0)\sin 2\theta\sin 2\varphi\\&\qquad+b\cos\theta\sin\varphi)],\end{aligned}\quad(14)$$

$$\begin{aligned}N_\varphi=A[(&-8d(J_1/J_0)\sin\theta\\&-4f(J_1/J_0)\sin\theta\cos^2\varphi-a\cos\varphi)\\&+j(-8c(J_1/J_0)\sin\theta\\&-4e(J_1/J_0)\sin\theta\cos^2\varphi+b\cos\varphi)].\end{aligned}\quad(15)$$

## FIELD PATTERN OF SHORT-CIRCUITED CIRCULAR LOOP

The radiation intensity[14] is

$$\Phi=\frac{15\pi}{\lambda^2}(N_\theta N_\theta{}^*+N_\varphi N_\varphi{}^*),\quad(16)$$

where asterisk denotes conjugate quantity. Therefore, using formulas (14) and (15)

$$\begin{aligned}\Phi=\frac{15\pi|A|^2}{\lambda^2}\Bigg[&\left(f\frac{J_1}{J_0}\sin 2\theta\sin 2\varphi+a\cos\theta\sin\varphi\right)^2\\&+\left(e\frac{J_1}{J_0}\sin 2\theta\sin 2\varphi+b\cos\theta\sin\varphi\right)^2\\&+\left(8d\frac{J_1}{J_0}\sin\theta+4f\frac{J_1}{J_0}\sin\theta\cos^2\varphi+a\cos\varphi\right)^2\\&+\left(8c\frac{J_1}{J_0}\sin\theta+4e\frac{J_1}{J_0}\sin\theta\cos^2\varphi-b\cos\varphi\right)^2\Bigg].\end{aligned}\quad(17)$$

In practical antennas $d\gg f$ and $c\gg e$; therefore, neglecting the terms of the order of $f$ and $e$ in the last two round brackets:

$$\begin{aligned}\Phi=\frac{15\pi|A|^2}{\lambda^2}\Bigg[&\left(f\frac{J_1}{J_0}\sin 2\theta\sin 2\varphi+a\cos\theta\sin\varphi\right)^2\\&+\left(e\frac{J_1}{J_0}\sin 2\theta\sin 2\varphi+b\cos\theta\sin\varphi\right)^2\\&+\left(8d\frac{J_1}{J_0}\sin\theta+a\cos\varphi\right)^2\\&+\left(8c\frac{J_1}{J_0}\sin\theta-b\cos\varphi\right)^2\Bigg].\end{aligned}\quad(18)$$

The field intensity[15] is $|E|=(240\pi\Phi)^{\frac{1}{2}}/r$. Using

---

[14] S. A. Schelkunoff, reference 10, p. 235.

[15] S. A. Schelkunoff, reference 10, p. 333.

formula (18), the horizontal field pattern is:

$$|E_h| = B\left[\left(8d\frac{J_1}{J_0}+a\cos\varphi\right)^2+\left(8c\frac{J_1}{J_0}-b\cos\varphi\right)^2\right]^{\frac{1}{2}}, \tag{19}$$

where $B=60\pi|A|/r\lambda$.

From the definition of the coefficients $a$, $b$, $c$, and $d$, it follows that: $a>0$, $b>0$, $c<0$, $d>0$. Therefore, the maximum field intensity in the horizontal plane is for $\varphi=0$

$$|E_{h,\max}| = B\left[\left(8d\frac{J_1}{J_0}+a\right)^2+\left(8c\frac{J_1}{J_0}-b\right)^2\right]^{\frac{1}{2}}. \tag{20}$$

The field intensity in the horizontal plane for $\varphi=\pi$ is

$$|E_{h,\pi}| = B[(8d(J_1/J_0)-a)^2 + (8c(J_1/J_0)+b)^2]^{\frac{1}{2}}. \tag{21}$$

The ratio of field intensities in $\varphi=\pi$ and $\varphi=0$ directions is, therefore,

$$|E_{h,\pi}|/|E_{h,\max}| = \left[\frac{(8d(J_1/J_0)-a)^2+(8c(J_1/J_0)+b)^2}{(8d(J_1/J_0)+a)^2+(8c(J_1/J_0)-b)^2}\right]^{\frac{1}{2}}. \tag{22}$$

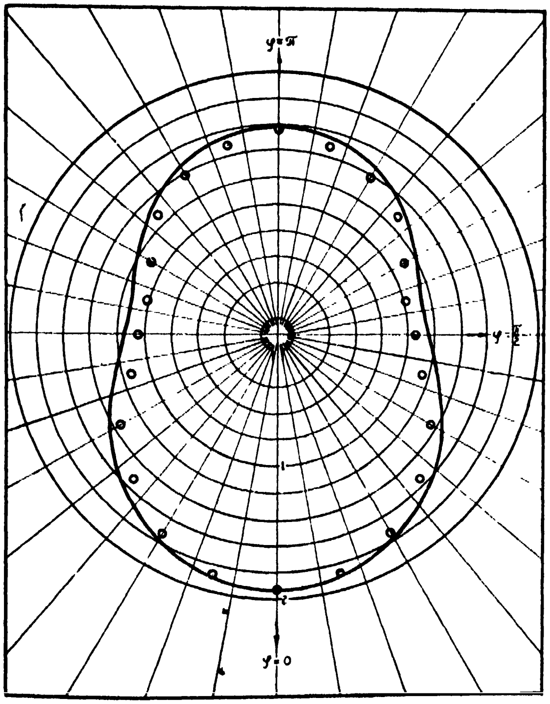


FIG. 4. Horizontal field pattern for $\alpha\bar{r}=0.04$ and $\beta\bar{r}=0.5$. Encircled points were obtained experimentally.

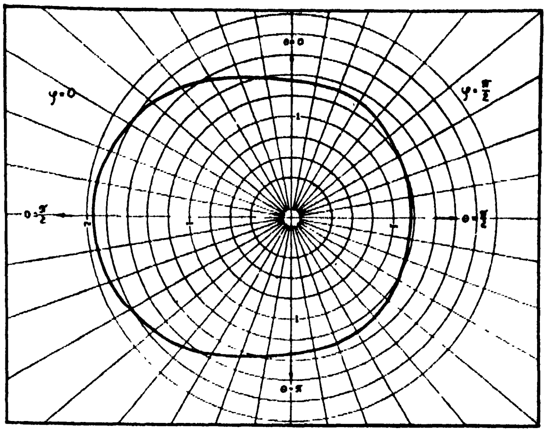


FIG. 5. Vertical field pattern in ($\varphi=0$)-(left) and ($\varphi=\pi/2$)-(right) planes for $\alpha\bar{r}=0.04$ and $\beta\bar{r}=0.5$.

From the last formula and the definition of coefficients $a$, $b$, $c$, $d$, it follows that the ratio of field intensities in $\varphi=\pi$ and $\varphi=0$ directions is a function of $\alpha\bar{r}$ and $\beta\bar{r}$. Figure 2 shows this ratio *versus* $\alpha\bar{r}$ for the important practical case of $\beta\bar{r}=0.5$.

The field intensity in the horizontal plane for $\varphi=\pi/2$ (or $3\pi/2$) is

$$|E_{h,\pi/2}| = B\left[\left(8d\frac{J_1}{J_0}\right)^2+\left(8c\frac{J_1}{J_0}\right)^2\right]^{\frac{1}{2}}. \tag{23}$$

The ratio of field intensities in $\varphi=\pi/2$ and $\varphi=0$ directions is

$$|E_{h,\pi/2}|/|E_{h,\max}| = \left[\frac{(8d(J_1/J_0))^2+(8c(J_1/J_0))^2}{(8d(J_1/J_0)+a)^2+(8c(J_1/J_0)-b)^2}\right]^{\frac{1}{2}}. \tag{24}$$

This ratio is again a function of $\alpha\bar{r}$ and $\beta\bar{r}$. Figure 3 shows this ratio *versus* $\alpha\bar{r}$ for $\beta\bar{r}=0.5$. Figure 4 is the plot of horizontal field pattern according to formula (18) for $\alpha\bar{r}=0.04$, and $\beta\bar{r}=0.5$. The encircled points represent the experimental pattern plotted to the same scale, and having the same value of $|E_{h,\pi}|/|E_{h,\max}|$ as the theoretical one.

It can be concluded that the horizontal field pattern of the circular short-circuited loop has the following characteristics:

1. The pattern is symmetrical about the $(0, \pi)$-direction.
2. There is the maximum field intensity in ($\varphi=0$)-direction (feeder end).

The vertical field pattern is

$$|E_v| = B[(f(J_1/J_0)\sin 2\theta \sin 2\varphi + a\cos\theta \sin\varphi)^2 + (e(J_1/J_0)\sin 2\theta \sin 2\varphi + b\cos\theta\sin\varphi)^2 + (8d(J_1/J_0)\sin\theta + a\cos\varphi)^2 + (8c(J_1/J_0)\sin\theta - b\cos\varphi)^2]^{\frac{1}{2}}. \quad (25)$$

In any $\varphi$-direction, the field intensity has a maximum value for $\theta=\pi/2$, that is in the horizontal plane.

The zenithal value of the field intensity ($\theta=0$) is

$$|E_{v,z}| = B(a^2+b^2)^{\frac{1}{2}}, \quad (26)$$

and is independent of $\varphi$.

Figure 5 is the plot of vertical field pattern according to formula (24) for $\alpha\bar{r}=0.04$ and $\beta\bar{r}=0.5$, for $\varphi=0$ and $\varphi=\pi/2$.

## RADIUS OF EQUIVALENT CIRCULAR HORIZONTAL FIELD PATTERN

The field intensity of the equivalent circular horizontal field pattern can be found from the formula

$$\pi|E_{h,\mathrm{Av}}|^2 = \tfrac{1}{2}\int_0^{2\pi} |E_h|^2 d\varphi. \quad (27)$$

Using formula (19) and after the elementary integration (Appendix 4)

$$|E_{h,\mathrm{Av}}| = B[64(J_1/J_0)^2(d^2+c^2)+0.5(a^2+b^2)]^{\frac{1}{2}}. \quad (28)$$

## POWER GAIN OF SHORT-CIRCUITED CIRCULAR LOOP

The radiation intensity from formula (18), when reduced to the unity maximum, is

$$\Phi^1 = \frac{\Phi}{\Phi_{\theta=\pi/2,\,\varphi=0}} = \frac{1}{64(J_1/J_0)^2(d^2+c^2)+16(J_1/J_0)(ad-bc)+(a^2+b^2)}[(J_1/J_0)^2(e^2+f^2)\sin^2 2\theta\sin^2 2\varphi + 2(J_1/J_0)(af+be)\sin 2\theta\cos\theta\sin 2\varphi\sin\varphi + (a^2+b^2)\cos^2\theta\sin^2\varphi + 64(J_1/J_0)^2(d^2+c^2)\sin^2\theta + 16(J_1/J_0)(ad-bc)\cos\varphi\sin\theta + (a^2+b^2)\cos^2\varphi]. \quad (29)$$

The radiated power[16] is

$$P = \int_{\theta=0}^{\theta=\pi}\int_{\varphi=0}^{\varphi=2\pi} \Phi^1 \sin\theta d\theta d\varphi. \quad (30)$$

Therefore (Appendix 5):

$$P = \frac{(8\pi/3)\{(J_1/J_0)^2[0.4(e^2+f^2)+64(d^2+c^2)]+(a^2+b^2)\}}{64(J_1/J_0)^2(d^2+c^2)+16(J_1/J_0)(ad-bc)+(a^2+b^2)}. \quad (31)$$

If short dipole is used as a reference,[17] the power gain is

$$(PG) = \frac{(8\pi/3)}{P} = \frac{64(J_1/J_0)^2(d^2+c^2)+16(J_1/J_0)(ad-bc)+(a^2+b^2)}{(J_1/J_0)^2[0.4(e^2+f^2)+64(d^2+c^2)]+(a^2+b^2)}. \quad (32)$$

## AVERAGE POWER GAIN

In f-m broadcasting, it is customary to base the gain on the equivalent circular horizontal field pattern. Since the average radiation intensity in the horizontal plane is

$$\Phi_{\mathrm{Av}} = r^2|E_{h,\mathrm{Av}}|^2/240\pi = (r^2B^2/240\pi)[64(J_1/J_0)^2(d^2+c^2) + 0.5(a^2+b^2)], \quad (33)$$

and the maximum radiation intensity in the horizontal plane is

$$\Phi_{\max} = \Phi_{\theta=\pi/2,\,\varphi=0} = \frac{r^2B^2}{240\pi}\left[\left(8d\frac{J_1}{J_0}+a\right)^2+\left(8c\frac{J_1}{J_0}-b\right)^2\right]; \quad (34)$$

therefore, the average power gain is

$$(PG)_{\mathrm{Av}} = \frac{\Phi_{\mathrm{Av}}}{\Phi_{\max}}(PG) = \frac{64(J_1/J_0)^2(d^2+c^2)+0.5(a^2+b^2)}{(J_1/J_0)^2[0.4(e^2+f^2)+64(d^2+c^2)]+(a^2+b^2)}. \quad (35)$$

---

[16] S. A. Schelkunoff, reference 10, p. 333.

[17] S. A. Schelkunoff, reference 10, p. 337.

The graph of $(PG)_{\mathrm{Av}}$ *versus* $\beta\bar{r}$ for $\alpha\bar{r}=0.04$ is shown in Fig. 6. (Since $c\gg d$ and $a\gg b$ in the practical range of $\alpha\bar{r}$, the gain computed for $\alpha\bar{r}=0$ may be applied also to the general case when $\alpha\bar{r}\neq 0$.)

## TRANSMISSION LINE EQUIVALENT TO THE CIRCULAR LOOP WITH NON-UNIFORM CURRENT DISTRIBUTION

The radiated power $P$ is given by formula (31). If the maximum current (at $\varphi=\pi$) is $I$, the radiation resistance $R_r$ is

$$R_r=\frac{2P}{I^2}=\frac{16\pi}{3I^2}\frac{(J_1/J_0)^2[0.4(e^2+f^2)+64(d^2+c^2)]+(a^2+b^2)}{64(J_1/J_0)^2(d^2+c^2)+16(J_1/J_0)(ad-bc)+(a^2+b^2)}. \tag{36}$$

Since $P$ in formula (31) was calculated for maximum radiation intensity equal to unity, therefore, from formula (18):

$$I^2=1\Big/\frac{15\pi|A|^2}{\lambda^2}\left[\left(8d\frac{J_1}{J_0}+a\right)^2+\left(8c\frac{J_1}{J_0}-b\right)^2\right]. \tag{37}$$

$$R_r=\frac{80\pi^2|A|^2}{\lambda^2}\left\{\left(\frac{J_1}{J_0}\right)^2[0.4(e^2+f^2)+64(d^2+c^2)]+(a^2+b^2)\right\}. \tag{38}$$

To calculate the attenuation constant $\alpha$ of the equivalent transmission line, let its resistance per unit length be $\rho$, and the current

$$I_{\bar{\varphi}}=I\cos\beta\bar{r}(\pi-\bar{\varphi}).$$

The power dissipated in the entire line is

$$\iota=\int^{\pi}\tfrac{1}{2}I_{\bar{\varphi}}^2\rho\bar{r}d\bar{\varphi}$$

$$=\tfrac{1}{4}I^2\rho\bar{r}\pi[1+(\sin 2\beta\bar{r}\pi/2\beta\bar{r}\pi)]. \tag{39}$$

From the equality of radiated and dissipated power:

$$\pi\rho\bar{r}=\frac{2R_r}{1+(\sin 2\beta\bar{r}\pi/2\beta\bar{r}\pi)}. \tag{40}$$

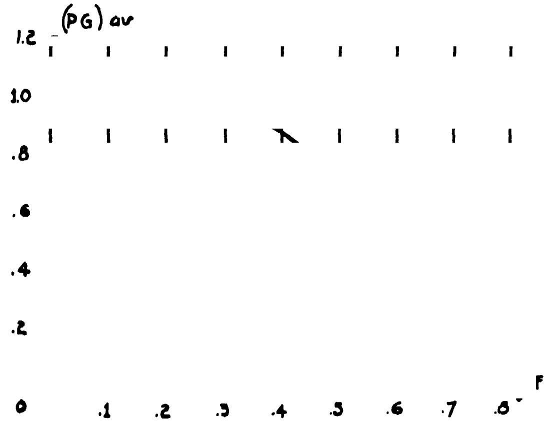


FIG. 6. Average power gain *versus* $\beta\bar{r}$ for $\alpha\bar{r}=0$.

On the other hand[18]

$$\alpha\bar{r}\pi=\frac{\pi\rho\bar{r}}{2Z_0}=\frac{R_r}{Z_0[1+(\sin 2\beta\bar{r}\pi/2\beta\bar{r}\pi)]}, \tag{41}$$

where $Z_0$ is the characteristic impedance of equivalent transmission line. According to Libby[19]

$$Z_0=276\log_{10}(\bar{r}/d), \tag{42}$$

where $\bar{r}$ is the mean loop radius, and $d$ the loop conductor diameter.

## ACKNOWLEDGMENTS

The author wishes to thank Mr. E. S. Kelsey for his help during the preparation of this paper, and Mr. J. C. Murray for computing aid.

## APPENDIX I

### Bessel Expansion of sin*t* and cos*t*

Since $\sin\theta\cos(\varphi-\bar{\varphi})=\frac{1}{2}(\sin\delta+\sin\epsilon)$, where $\delta=\theta+\varphi-\bar{\varphi}$, $\epsilon=\theta-\varphi+\bar{\varphi}$. Therefore,

$$\sin t=\sin\left[\frac{\beta\bar{r}}{2}\sin\delta\right]\cos\left[\frac{\beta\bar{r}}{2}\sin\epsilon\right]+\cos\left[\frac{\beta\bar{r}}{2}\sin\delta\right]\sin\left[\frac{\beta\bar{r}}{2}\sin\epsilon\right].$$

Using first terms of Bessel expansion[12] for sin and cos, $\sin t\leqq 4J_0J_1\sin\theta\cos(\varphi-\bar{\varphi})$, where $J_0\equiv J_0(\beta\bar{r}/2)$, $J_1\equiv J_1(\beta\bar{r}/2)$. Similarly,

$$\cos t=\cos[(\beta\bar{r}/2)\sin\delta]\cos[(\beta\bar{r}/2)\sin\epsilon]-\sin[(\beta\bar{r}/2)\sin\delta]\sin[(\beta\bar{r}/2)\sin\epsilon].$$

---

[18] S. A. Schelkunoff, reference 10, p. 196.

Using again first terms of Bessel expansion for sin and cos, $\cos t \cong J_0^2$, since $J_1^2 \ll J_0^2$.

## APPENDIX II

## Definite Integrals Appearing in Formulas (6)

Using Dwight's "Tables of Integrals" and elementary trigonometric identities:

$$\int_0^{\pi} \cosh\gamma\hat{r}\,\sin\bar{\varphi}\cos(\varphi-\bar{\varphi})d\bar{\varphi} =\frac{1}{4+\gamma^2\hat{r}^2}\left[(1-\cos j\gamma\hat{r}\pi)\cos\varphi+\frac{2\sin j\gamma\hat{r}\pi}{j\gamma\hat{r}}\sin\varphi\right], \tag{A2-1}$$

$$\int_0^{\pi} \sinh\gamma\hat{r}\bar{\varphi}\,\sin\bar{\varphi}\cos(\varphi-\bar{\varphi})d\bar{\varphi} =\frac{j}{4+\gamma^2\hat{r}^2}\left[\sin j\gamma\hat{r}\pi\cos\varphi+\frac{2(\cos j\gamma\hat{r}\pi-1)}{j\gamma\hat{r}}\sin\varphi\right], \tag{A2-2}$$

$$\int_0^{\pi} \cosh\gamma\hat{r}\bar{\varphi}\,\sin\bar{\varphi}d\bar{\varphi}=\frac{1+\cos j\gamma\hat{r}\pi}{1+\gamma^2\hat{r}^2}, \tag{A2-3}$$

$$\int_0^{\pi} \sinh\gamma\hat{r}\bar{\varphi}\,\sin\bar{\varphi}d\bar{\varphi}=-j\frac{\sin j\gamma\hat{r}\pi}{1+\gamma^2\hat{r}^2}. \tag{A2-4}$$

Therefore, after reduction, formula (6) yields

$$N_x=I\hat{r}J_0\left[-jJ_1\sin\theta\frac{16\sin j\gamma\hat{r}\pi}{j\gamma\hat{r}(4+\gamma^2\hat{r}^2)}\sin\varphi\right]. \tag{A2-5}$$

Similarly

$$\int_0^{\pi} \cosh\gamma\hat{r}\varphi\,\cos\bar{\varphi}\cos(\varphi-\bar{\varphi})d\varphi =\frac{1}{4+\gamma^2\hat{r}^2}\left[\frac{(2+\gamma^2\hat{r}^2)\sin j\gamma\hat{r}\pi}{j\gamma\hat{r}}\cos\varphi +(1-\cos j\gamma\hat{r}\pi)\sin\varphi\right], \tag{A2-6}$$

$$\int_0^{\pi} \sinh\gamma\hat{r}\bar{\varphi}\,\cos\bar{\varphi}\cos(\varphi-\bar{\varphi})d\bar{\varphi} =\frac{j}{4+\gamma^2\hat{r}^2}\left[\frac{(2+\gamma^2\hat{r}^2)(\cos j\gamma\hat{r}\pi-1)}{j\gamma\hat{r}}\cos\varphi +\sin j\gamma\hat{r}\pi\sin\varphi\right], \tag{A2-7}$$

$$\int_0^{\pi} \cosh\gamma\hat{r}\bar{\varphi}\,\cos\bar{\varphi}d\bar{\varphi}=\frac{j\gamma\hat{r}\sin j\gamma\hat{r}\pi}{1+\gamma^2\hat{r}^2}, \tag{A2-8}$$

$$\int_0^{\pi} \sinh\gamma\hat{r}\bar{\varphi}\,\cos\bar{\varphi}d\bar{\varphi}=j\frac{j\gamma\hat{r}(1+\cos j\gamma\hat{r}\pi)}{1+\gamma^2\hat{r}^2}. \tag{A2-9}$$

Therefore, after reduction, the second formula (6) yields

$$N_y=I\hat{r}J_0\left[jJ_1\sin\theta\frac{8(2+\gamma^2\hat{r}^2)\sin j\gamma\hat{r}\pi}{j\gamma\hat{r}(4+\gamma^2\hat{r}^2)}\cos\varphi -J_0\frac{2j\gamma\hat{r}\sin j\gamma\hat{r}\pi}{1+\gamma^2\hat{r}^2}\right]. \tag{A2-10}$$

## APPENDIX III

## Approximations to Formula (11)

Write

$$\gamma\hat{r}=\alpha\hat{r}+j\beta\hat{r}=j\beta\hat{r}[1-j(\alpha/\beta)].$$

Since $\alpha \ll \beta$

$$(\gamma\hat{r})^2\cong-(\beta\hat{r})^2\left[1-j\frac{2\alpha\hat{r}\beta\hat{r}}{(\beta\hat{r})^2}\right].$$

Therefore,

$$\frac{1}{1+\gamma^2\hat{r}^2}\cong\frac{1}{1-(\beta\hat{r})^2}-j\frac{2\alpha\hat{r}\beta\hat{r}}{1-2(\beta\hat{r})^2}=a-jb,$$

$$\frac{1}{4+\gamma^2\hat{r}^2}\cong\frac{1}{4-(\beta\hat{r})^2}-j\frac{2\alpha\hat{r}\beta\hat{r}}{16-8(\beta\hat{r})^2}=e-jf,$$

$$\frac{1}{\gamma^2\hat{r}^2(4+\gamma^2\hat{r}^2)}\cong-\frac{1}{(\beta\hat{r})^4[16-8(\beta\hat{r})^2]}-j\frac{2\alpha\hat{r}\beta\hat{r}}{4(\beta\hat{r})^4}=c-jd.$$

## APPENDIX IV

## Integral in Formula for Radius of Equivalent Circular Horizontal Pattern

From formula (19)

$$|E_h|^2=B^2[64(J_1/J_0)^2(d^2+c^2)+16(J_1/J_0)(ad-bc)\cos\varphi +(a^2+b^2)\cos^2\varphi].$$

Therefore, after the elementary integration

$$\int_0^{2\pi}|E_h|^2d\varphi=\pi B^2[128(J_1/J_0)^2(d^2+c^2)+(a^2+b^2)].$$

## APPENDIX V

## Integration of Formula (30)

Using Dwight's "Tables of Integrals," the integrals appearing in formula (30) are:

$$\int_0^{\pi}\sin^2 2\theta\sin\theta d\theta\int_0^{2\pi}\sin^2 2\varphi d\varphi=\frac{16}{15}\pi,$$

$$\int_0^{\pi}\sin 2\theta\sin\theta\cos\theta d\theta\int_0^{2\pi}\sin 2\varphi\sin\varphi d\varphi=0,$$

$$\int_0^{\pi}\sin\theta\cos^2\theta d\theta\int_0^{2\pi}\sin^2\varphi d\varphi=\frac{2}{3}\pi,$$

$$\int_0^{\pi}\sin^3\theta d\theta\int_0^{2\pi}d\varphi=\frac{8}{3}\pi,$$

$$\int_0^{\pi}\sin^2\theta d\theta\int_0^{2\pi}\cos\varphi d\varphi=0,$$

$$\int_0^{\pi}\sin\theta d\theta\int_0^{2\pi}\cos^2\varphi d\varphi=2\pi.$$

# High Speed Compression Tests on Copper

M. Greenfield, *North American Aviation, Los Angeles, California*

AND

E. T. Habib, *David Taylor Model Basin, Carderock, Maryland*

(Received December 19, 1946)

Dynamic shortening of one-half inch long copper cylinders is achieved by striking them with a hardened steel projectile at high velocities. The average strain rate was about 1200 per second. Energy per unit volume absorbed by the copper is plotted against strain. A true stress-logarithmic strain curve is computed. This curve is compared with a similar curve derived from high speed tests on copper in tension.

## INTRODUCTION

THIS work was undertaken originally to provide a dynamic calibration for copper cylinders. These cylinders are used in mechanical gauges which are designed to measure the pressure in the water originating from an under-water explosion. The copper is deformed dynamically in compression in such gauges. The experimental arrangement described below was designed so that the copper cylinders were subjected to the same kind of forces as they were when used in the gauges. However, it was thought that additional information could be obtained from the data which would be of wider interest.

These data have been put in the form of stress *versus* strain curves. Usually stress is defined as the force acting on the specimen divided by the original cross-sectional area of the test portion, i.e.,

$$\sigma_c = F/A_0$$

where $A_0$ is the original cross-sectional area. This stress, $\sigma_c$, is also called the engineering stress. The strain is frequently defined as the change in length per unit original length, i.e.,

$$\epsilon_c = \Delta l/l_0$$

where $l_0$ is the original length.

While some work has been done by Nadai and Manjoine on high speed tension tests of metals,[1] very little has been reported on high speed compression tests such as those described below. It was thought useful to compare the properties of copper in tension with those in compression when the forces are applied rapidly. The parameter chosen to describe the rapidity of the test is the time rate of strain $\dot{\epsilon} = d\epsilon_c/dt$.

For purposes of comparison of the properties of a metal in compression with those in tension, the choice of variables as defined above is poor. In tension tests the cross-sectional area of a specimen is diminished, making the true stress based on actual area greater than the engineering stress. In compression tests this area is increased, making the true stress less than the conventional stress. There is a similar difficulty in the comparison of strains for tension and compression tests. For these reasons it was decided to use the so-called true stress and logarithmic strain. For compression tests these quantities may be defined as follows:

$$\sigma = F/A$$

where $\sigma$ is the true stress, $F$ is the applied force, and $A$ is the instantaneous cross-sectional area of the specimen.

$$\epsilon = \ln A_0/A,$$

where $\epsilon$ is the logarithmic strain, and $A_0$ is the original cross-sectional area of the specimen. The logarithm is in terms of the base $e$.

These quantities $\sigma$ and $\epsilon$ are simply related to the conventional ones $\sigma_c$ and $\epsilon_c$, as follows:

$$\sigma = \sigma_c(1-\epsilon_c), \qquad (1)$$

$$\epsilon = \ln 1/(1-\epsilon_c). \qquad (2)$$

These relations are derived[2] by making the usual assumption for the plastic flow of metals

---

[1] M. Manjoine and A. Nadai, Proc. A.S.T.M. 40, 822 (1940).

[2] A. Nadai, App. Phys. 8, 205 (1937).

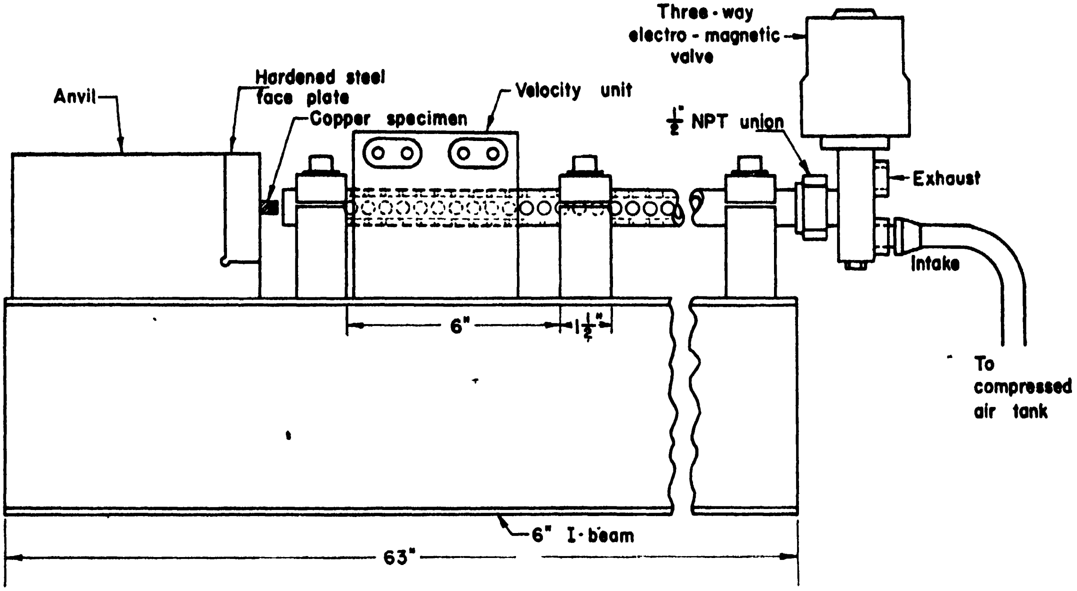


FIG. 1. Air gun for compressing copper cylinders.

that the volume of any element of metal remains unchanged.

## APPARATUS

The test equipment consists of an air gun for blowing a hardened steel piston against the soft copper specimen and auxiliary equipment for measuring the energy of the piston (see Figs. 1 and 2). The gun barrel is a cylindrical tube of steel 55″ long with an outside diameter of $\frac{15}{16}$ inch and an inner diameter of $\frac{1}{2}$ inch. At the muzzle end of the barrel, spaced $\frac{1}{8}$ inch apart, one hundred eighty $\frac{1}{4}$″ holes are drilled in the barrel.

The gun barrel is laid in a V-shaped groove cut into a set of steel supports. Similarly grooved hold-down plates, bolted to the supports, keep the barrel in position. These supports are in turn bolted to a 6-inch I-beam, about 63 inches long. At the muzzle end of the barrel a steel anvil, with a hardened steel face plate, is bolted to the I-beam. The test specimen is placed against this steel plate. The face of the hardened steel plate is very nearly perpendicular to the line of travel of the piston. This insures that the faces of the copper specimen are almost parallel after deformation. The variation in readings at different points on the faces of the specimen is thus kept small.

The propulsive force for the piston is supplied by a compressed air system. Air from a tank at high pressure is led into a smaller tank through a regulating valve. This regulator enables the small tank to be filled to any pressure up to 400 lb. per square inch, although 100 lb./in.$^2$ was the maximum used. A 3-way quick-acting solenoid valve controls the flow of air from the small reservoir to the gun barrel. The time of opening and closing of the valve is determined by a time-delay relay. The sequence of events is the following. A double-pole, double-throw switch is thrown. One side of the switch closes a circuit allowing 220-volt a.c. to flow through the solenoid; thus energized, the solenoid opens the intake valve and allows air to blow against the piston. The piston then moves down the barrel, strikes the specimen and rebounds. Simultaneously with the closure of the first side of the switch the second side closes a circuit in the time-delay unit. After a predetermined time delay a relay is thrown which opens the solenoid circuit. This closes the intake valve and, at the same time, opens the exhaust valve, allowing the rebounding piston to compress and force air out through the exhaust; in so doing, the piston brings itself to a halt.

By actual test, the shortest time for which the time-delay unit could be set to operate properly was about 0.01 second. This setting was used at the fastest velocities and worked very well. In the other direction, the time-delay unit may be set to operate for as much as several seconds. The purpose in keeping the valve open for only a short time is to keep the excess of air flowing through the barrel to a minimum.

The test specimens are made from oxygen-free high conductivity copper rod. Each specimen is made so that the final length is 0.5000 inch ±0.0005 inch and the diameter is 0.3255 inch ±0.0005 inch. They are then annealed in an electrically heated furnace in an atmosphere of hydrogen at a temperature of 950° Fahrenheit. The temperature is maintained for 2½ to 3 hours and the pellets allowed to cool in the furnace. The flow of hydrogen is maintained until the temperature is below 250° Fahrenheit.

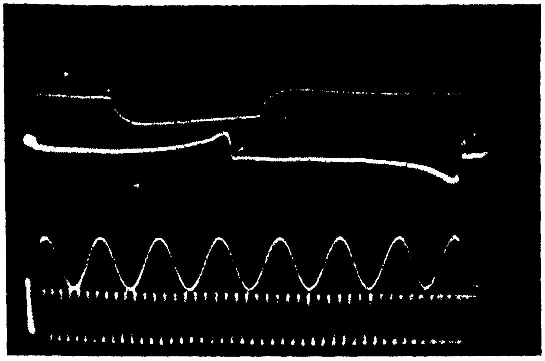


FIG. 3. A typical record obtained by the interruption of the light beam by the moving piston. At (*A*), the piston first enters the light beam; at (*B*), it is leaving the light beam. (*C*) and (*D*) represent a similar situation after the piston has rebounded. The lower oscillogram is the timing calibration for *AB* and *CD*.

The velocity of the piston is measured by noting the time interval between the breaking and making of a narrow beam of light which is incident upon a photo-cell (see Fig. 2). As the piston breaks the first light beam, a voltage change is induced in the photo-cell. This voltage is then amplified and used to trip a sweep generator. The sweep generator controls the horizontal movement of an electron beam across the face of a cathode-ray oscillograph. When the electron spot has completed about a third of its journey, the piston interrupts the second beam of light, producing a voltage impulse

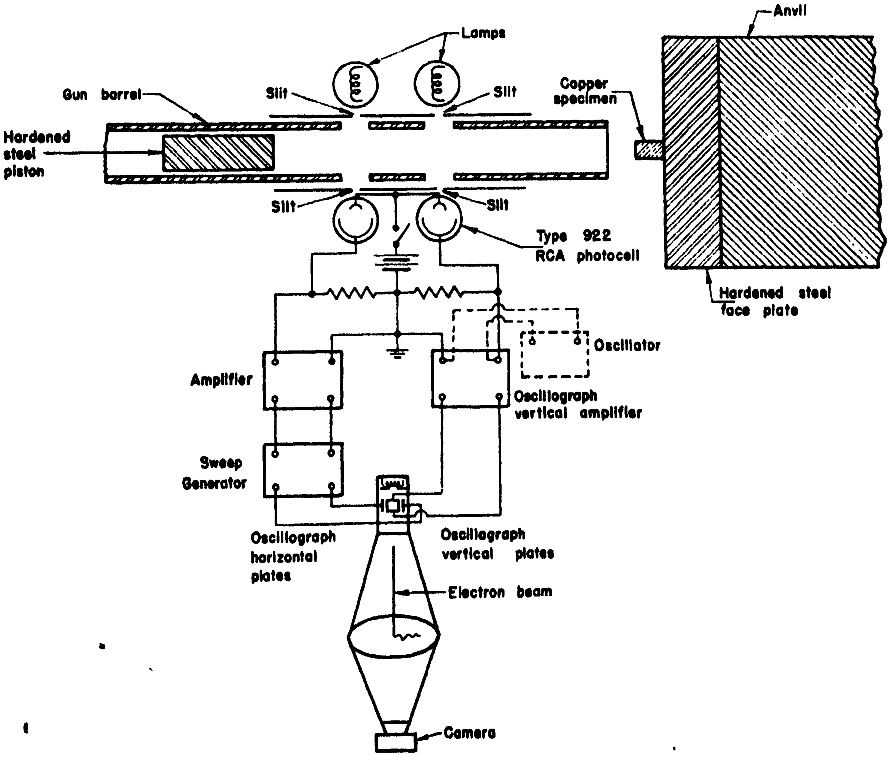


FIG. 2. Apparatus for measuring velocity of piston.

which is led into the vertical amplifier of the oscillograph. The spot then jumps up. When the piston has passed the light beam, the spot jumps back (see Fig. 3). After the piston strikes the copper, it rebounds; this velocity is recorded on the return trace of the electron beam. The distance between jumps is related to time by comparing it with a sine wave of known frequency. This sine wave is taken from an audio-frequency oscillator. To insure accurate time measurements, the output of the oscillator is beat against that of a 1000-cycle tuning-fork oscillator before the timing oscillogram is made. Knowing the piston length and its time to pass by the light beam, the velocity of the piston may be calculated. The energy of deformation is then computed from the relation K.E. $= \frac{1}{2}m(v_1^2 - v_2^2)$ where $m$ is the mass of the piston, $v_1$ is its velocity before striking the specimen, and $v_2$ is its velocity of rebound. The oscilloscope tracings were photographed on 35-mm Super XX film in a camera equipped with an $f$:1.9 lens.

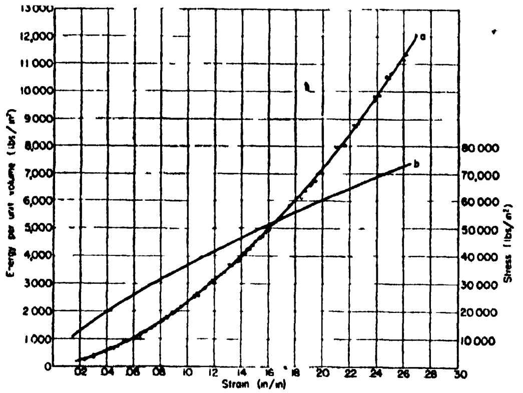


FIG. 4. Energy per unit volume-strain curve and stress-strain curve for copper cylinders in compression. Curve (a) shows dependence of energy absorbed per unit volume on the strain in compression in copper cylinders. Curve (b) is a stress-strain curve for copper derived from curve (a) by numerical differentiation. The rate of strain is approximately 1200 per second.

## PRECISION OF MEASUREMENTS

Before being subjected to impact, each cylinder was measured with a micrometer. One reading was made at the center, four at intervals around the circumference. The average of these was taken as the length of the pellet before deformation. The averages of two sets of reading made in this way repeated to within 0.0001″. After impact, the pellet was again measured at several points around the circumference and an average computed. (The center reading was discarded because it was consistently higher than the average.) The averages of two sets of readings in this case always agreed to within 0.0002″. The maximum error possible from this measurement is then 0.0003″. When the deformation is greater than 0.03″, this error is less than 1 percent.

Measurements on the film are made with a traveling microscope. Measurements with this instrument can be made to a precision of 0.001 millimeter. The line thickness on the film was about 0.1 millimeter. The distance between jumps on the film was of the order of 5 millimeters. Measurements were made to the center of each line. Three or more readings were taken at each point and averaged. The averages could be repeated to within 0.01 millimeter. The maximum error possible in this measurement then is about 0.2 percent.

Measurements on the sine wave were made at the same portion of the film as the velocity record occurred. This was essential because the sweep was not exactly linear, especially at the edges of the scope face. However in the middle half of the oscillogram, the linearity was within 1 percent. The oscillator frequency was always made high enough so that about five cycles covered the same distance as the velocity record (see Fig. 3). Errors due to non-linearity of the sweep were thus reduced. A timing oscillogram was made before and after every few records. Velocities were computed using both timing records and an average taken. Velocities based on different timing oscillograms varied from the average by less than 0.5 percent.

The total maximum random error is thus about 2 percent.

## STRAIN RATE ($\dot{\epsilon}$)

Strain rate is defined as the time rate of change of the strain. Since the difference between the conventional strain and the logarithmic strain is small here, of order 10 percent, and since the dependence of stress on strain rate turns out to be a very slowly changing function (roughly multiplying the strain rate by ten means an increase of 4 or 5 percent in stress), it is sufficiently accurate to use the time rate of change of the conventional strain for the strain rate. For the ½-inch long coppers used here, assuming the average velocity of the striking piston is one-half of the initial velocity with which it strikes the specimen, i.e., uniform deceleration, the average strain rate is numerically equal to the initial velocity expressed in inches/sec.

$$\text{Average } \dot{\epsilon} = \frac{\Delta l}{l_0}\cdot\frac{1}{T} = \frac{\Delta l}{T}\cdot\frac{1}{l_0}$$

$$= \frac{V_a}{l_0} = \frac{V_0}{2l_0} = V_0\frac{\text{in./sec.}}{\text{in.}},$$

where $\Delta l$ = final deformation, $l_0$ = original length = ½ inch, $T$ = total time of deformation-seconds, $v_0$ = initial velocity-inches/second, and $v_a$ = average velocity-inches/second.

The average strain rate is computed by assuming that the velocity with which the specimen contracts, i.e., the velocity of the striking piston, is a decreasing linear function of time. This assumption was checked by making two high speed photographs of the deformation of the copper cylinder. These showed that while the velocity of the striking piston is not linear, the true average velocity exceeded the assumed one by only about 15 percent. Because of the nature of the dependence of the stress on the strain rate, this method of computing the average strain rate is considered adequate.

## DATA

Figure 4, curve (4a), is a plot of energy/unit volume in lb./in.$^2$ against conventional strain in inches/inch. It represents the basic data of the dynamic testing in compression. A smooth curve was drawn through the points in such a manner that as many points lay above the curve as below it. One copper specimen was used for each point and 118 points were obtained. Two hardened steel pistons were employed; one weighed 0.05473 lb., the other 0.1633 lb. Through their use, the average strain rate at the upper and lower ends of the curve was kept within about 50 percent of 1200 l/sec.

Curve (4b) which shows conventional stress *versus* conventional strain was derived from curve (4a) by numerical differentiation. In order to compare these compression data with data in tension, curve (4b) was revised by applying formulas (1) and (2).

Figure 5 shows the revised stress-strain curves. In all cases the ordinate is true stress and the abscissa is logarithmic strain. Curves (a) and (b) in this figure show the relation between stress and strain for copper in compression. All the specimens used in obtaining curves (a) and (b) came from the same lot. The strain rates that apply to these curves are approximately 1200 l/sec. for curve (a) and $1.5\times10^{-3}$ l/sec. for curve (b). The latter curve was obtained by means of an ordinary static compression test. Curve (a) was based on data from the dynamic test and was obtained by the procedure described above. The stress required to produce a given compressional strain dynamically was always greater than that required to produce the same strain statically. This difference in stress increases as the strain increases. Comparison of curves (a) and (b) shows that the percentage difference varies from 17 percent to 29 percent for the interval of strain from 0.10 to 0.30.

It was thought useful to compare these results with those obtained in high speed tension tests of copper at room temperatures. Some work of this kind has been done by Manjoine and Nadai.[1] Curves (c) and (d) in Fig. 5 show the stress-strain curves obtained by them for copper. Their specimens were made of commercially pure copper and were 0.200″ in diameter and 1″ in length. After being machined the specimens were annealed by heating them in dry regenerated gas to 500°C. After that they were cooled to room temperature in 12 hours.

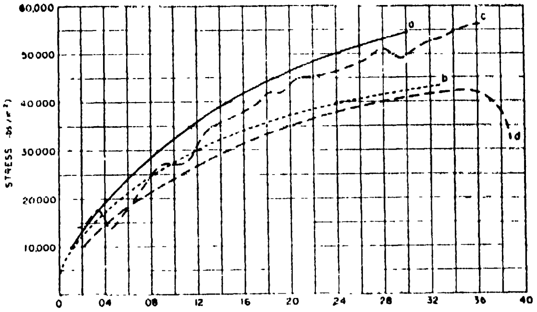


FIG. 5. Stress-strain curves for copper. The ordinate is true stress and the abscissa is logarithmic strain.

Curve (c) in Fig. 5 was obtained in a test of a single specimen subjected to a rate of strain of 900 per second. Curve (d) in Fig. 5 was obtained by means of a static tension test. The difference between the two static curves (b) and (d) is probably due to the fact that the copper used by the writers was not quite the same as that used by Nadai and Manjoine. Comparison of curves (c) and (d) shows that the stress required to produce a given strain in tension dynamically generally exceeds that required to produce the same strain statically. Some of the oscillations in curve (c) are probably not real but may be due to the method of loading the specimen and measuring that load. A more valid curve may possibly be obtained by drawing a smooth average line through the oscillations. If one does this, then the percentage difference between curves (c) and (d) varies from approximately 19 percent to 25 percent for the interval of strain from 0.10 to 0.30. It is interesting to note that the percentage increase in stress required to produce a given strain in copper dynamically rather than statically is much the same for both compression and tension. This comparison should be made when the rates of strain are approximately the same.

---

# $TM_{0,1}$ Mode in Circular Wave Guides with Two Coaxial Dielectrics

SIDNEY FRANKEL
*Federal Telecommunication Laboratories, Inc., New York, New York*

(Received February 3, 1947)

Field components for a transverse magnetic wave in a wave guide with two coaxial dielectrics are computed. A typical example is given to show the calculation of guide dimensions to reduce phase velocity to a preassigned value.

A UNIFORM circular wave guide may be conceived as a device that provides an interaction space between electrons and an electromagnetic field to transfer energy from one to the other. If electrons are injected axially in the guide and the field has an axial electric component, the important interaction is between the electrons and this axial component.

For unimpeded motion, the electrons must travel essentially in a vacuum and with a velocity necessarily less than that of light. On the other hand, the phase velocity of the electromagnetic field in an evacuated circular guide is always greater than the velocity of light. For effective interaction, this field must be slowed down, perhaps to the order of one-tenth or less of the velocity of light. This may be accomplished by surrounding the evacuated axial region of the guide with material of high dielectric constant. For simplicity, circular symmetry may be maintained.

An investigation of the propagation of an electromagnetic field with axial electric component in such a "loaded" guide has been made. Appropriate field equations, which take the necessary boundary conditions into account, are set up and solved. A relationship between the maximum value of axial electric component and transmitted power is given.

## 1. SOLUTION OF FIELD EQUATIONS

### 1.1 Differential Equations

Figure 1 shows a cross section of a wave guide. The section of the axial region is a circle of radius $a$, the constants of the medium being $\mu$ and $\epsilon_2$, the permeability and dielectric constants, respectively. The boundary of the guide is a perfect conductor of radius $b$. In the annular region between these two radii is a material of constants $\mu$ and $\epsilon_1$. It is assumed that $\epsilon_1 > \epsilon_2$ always.

Cylindrical coordinates $r$, $\phi$, and $x$ will be used with the $x$ axis coinciding with the axis of the guide and being positive in the direction toward a sink of power. For a transverse magnetic mode, take $H_x = 0$, and, in accordance with the conventional $TM_{0,1}$ mode for uniform dielectrics, specify additionally that $E_\phi = H_r = 0$. If a field satisfying these requirements exists in the guide under consideration, then the appropriate field equations in m.k.s. units for either dielectric are of the following type.

$$\left.\begin{aligned} -\partial H_\phi/\partial x &= j\omega\epsilon E_r,\\ (1/r)(\partial/\partial r)(rH_\phi) &= j\omega\epsilon E_x,\\ (1/r)(\partial E_x/\partial\phi) &= 0,\\ (\partial E_r/\partial x)-(\partial E_x/\partial r) &= -j\omega\mu H_\phi,\\ (1/r)(\partial E_r/\partial\phi) &= 0. \end{aligned}\right\} \quad (1)$$

The subscript has been omitted from $\epsilon$ to avoid restricting it to either medium.

According to (1), $E_r$ and $E_x$, therefore also $H_\phi$, are independent of $\phi$. The useful equations from (1) are

$$\left.\begin{aligned} \partial H_\phi/\partial x &= -j\omega\epsilon E_r,\\ (1/r)(\partial/\partial r)(rH_\phi) &= j\omega\epsilon E_x,\\ (\partial E_r/\partial x)-(\partial E_x/\partial r) &= -j\omega\mu H_\phi. \end{aligned}\right\} \quad (2)$$

Elimination of $E_r$ and $E_x$ among these equations leads to the following equation in $H_\phi$:

$$\frac{\partial^2 H_\phi}{\partial x^2}+\frac{\partial}{\partial r}\left[\frac{1}{r}\frac{\partial}{\partial r}(rH_\phi)\right]=-\omega^2\mu\epsilon H_\phi=-\frac{\omega^2}{c^2}H_\phi, \quad (3)$$

where $c$ is the velocity of propagation in a medium of unbounded extent with constants $\mu$ and $\epsilon$.

Assume

$$H_\phi=R(r)e^{-\gamma x}, \quad (4)$$

where $\gamma$ is the propagation constant in the $x$ direction and $R$ is a function of $r$ to be determined. Substitution of (4) in (3) yields

$$\frac{d}{dr}\left[\frac{1}{r}\frac{d}{dr}(rR)\right]+\alpha^2 R=0, \quad (5)$$

where

$$\alpha^2=\gamma^2+\omega^2/c^2. \quad (6)$$

Solutions of (5) are of the type

$$R=J_1(\alpha r),\ Y_1(\alpha r), \quad (7)$$

where $J_1$ and $Y_1$ are Bessel functions of the first and second kinds, respectively, and of the first order. If $Z_1(\alpha r)$ represents any linear combination of these functions, then

$$H_\phi=Z_1(\alpha r)e^{-\gamma x}. \quad (8)$$

For unattenuated transmission, $\gamma$ is a pure imaginary; let $\gamma=j\beta$, with $\beta$ real. Then

$$H_\phi=Z_1(\alpha r)e^{-j\beta x}, \quad \alpha=[(\omega/c)^2-\beta^2]^{\frac{1}{2}}. \quad (9)$$

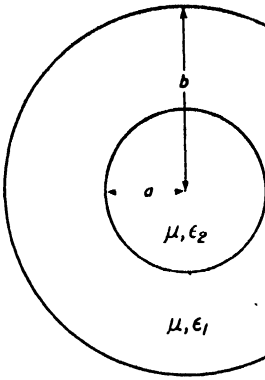


FIG. 1. Cross section of wave guide having two coaxial dielectrics.

From (2), we get immediately

$$\left.\begin{aligned} E_r &= (\beta/\omega\epsilon)Z_1(\alpha r)e^{-j\beta x},\\ E_x &= (\alpha/j\omega\epsilon)Z_0(\alpha r)e^{-j\beta x}, \end{aligned}\right\} \quad (10)$$

where $Z_0$ is a linear combination of Bessel functions of the first and second kind, zeroth order, and we have made use of the relation[1]

$$\frac{d}{dr}[rZ_1(\alpha r)]=\alpha rZ_0(\alpha r). \quad (11)$$

$J_0$ and $J_1$ are finite and continuous everywhere in the complex plane, but $Y_0$ and $Y_1$ have infinite discontinuities at $r=0$. These latter functions are therefore inadmissable in the axial region of Fig. 1. The solutions therefore run as follows.

In medium 1,

$$\left.\begin{aligned} H_{\phi 1} &= [A_1J_1(\alpha_1 r)+B_1Y_1(\alpha_1 r)]e^{-j\beta_1 x},\\ E_{r1} &= \frac{\beta_1}{\omega\epsilon_1}[A_1J_1(\alpha_1 r)+B_1Y_1(\alpha_1 r)]e^{-j\beta_1 x},\\ E_{x1} &= \frac{\alpha_1}{j\omega\epsilon_1}[A_1J_0(\alpha_1 r)+B_1Y_0(\alpha_1 r)]e^{-j\beta_1 x}. \end{aligned}\right\}$$

$$a\leqq r\leqq b. \quad (12a)$$

In medium 2,

$$\left.\begin{aligned} H_{\phi 2} &= A_2J_1(\alpha_2 r)e^{-j\beta_2 x},\\ E_{r2} &= (\beta_2/\omega\epsilon_2)A_2J_1(\alpha_2 r)e^{-j\beta_2 x},\\ E_{x2} &= (\alpha_2/j\omega\epsilon_2)A_2J_0(\alpha_2 r)e^{-j\beta_2 x}, \end{aligned}\right\} 0\leqq r\leqq a. \quad (12b)$$

---

[1] Jahnke and Emde, *Tables of Functions* (B. G. Teubner, Leipzig, 1938), third edition, p. 145.

$A_1$, $B_1$, and $A_2$ are constants to be determined from the boundary and power conditions, $\alpha_1$ and $\alpha_2$ are the propagation constants, and $\beta_1$ and $\beta_2$ are the phase constants, in media 1 and 2, respectively.

### 1.2 Boundary Conditions

Matching of tangential components of the electric field at the dielectric interface requires that

$$E_{x1}=E_{x2} \quad \text{at} \quad r=a$$

for all values of $x$; i.e.,

$$\frac{\alpha_1}{j\omega\epsilon_1}[A_1J_0(\alpha_1a)+B_1Y_0(\alpha_1a)]e^{-j\beta_1x} = -\frac{\alpha_2}{j\omega\epsilon_2}A_2J_0(\alpha_2a)e^{-j\beta_2x}. \tag{13}$$

This can be satisfied for all values of $x$ if, and only if, $\beta_1=\beta_2=\beta$. The phase constant (hence the velocity of propagation) is the same in both media, as might have been expected.

$$\alpha_1^2-\omega^2/c_1^2=-\beta^2=\alpha_2^2-\omega^2/c_2^2. \tag{14}$$

Assume the phase velocity $V_p$ to be intermediate between $c_1$ and $c_2$

$$c_1<V_p<c_2, \quad V_p=\omega/\beta. \tag{15}$$

This leads to the inequality

$$\alpha_2^2<0<\alpha_1^2, \tag{16}$$

whence $\alpha_1$ is real while $\alpha_2$ is a pure imaginary. The field in medium 2 therefore behaves according to Bessel functions of the first kind with pure imaginary argument. Curves for $J_0(jz)$ and $-jJ_1(jz)$, frequently designated as $I_0(z)$ and $I_1(z)$, respectively, have been published.[2] The function $I_0(z)$ is particularly interesting since it shows the radial variation of the axial electric field. This field is fairly uniform for sufficiently small $z$. For $z<0.6$, the variation is less than 10 percent; for $z<2$, the variation is less than about 2:1.

Next, for the tangential component of electric field to vanish at the conductor surface, it is required that $E_{x1}=0$ at $r=b$, i.e.,

$$A_1J_0(\alpha_1b)+B_1Y_0(\alpha_1b)=0. \tag{17}$$

Matching the normal components of dielectric flux at the dielectric interface requires that

$$\epsilon_1E_{r1}=\epsilon_2E_{r2} \quad \text{at} \quad r=a; \text{ i.e.,}$$
$$A_1J_1(\alpha_1a)+B_1Y_1(\alpha_1a)=A_2J_1(\alpha_2a). \tag{18}$$

Matching the tangential magnetic field at the dielectric interface yields no further information. Collecting and rearranging (13), (17), and (18), we have the set of linear homogeneous equations in $A_1$, $B_1$, and $A_2$:

$$\left.\begin{aligned} \frac{\alpha_1}{\epsilon_1}J_0(\alpha_1a)A_1+\frac{\alpha_1}{\epsilon_1}Y_0(\alpha_1a)B_1 -\frac{\alpha_2}{\epsilon_2}J_0(\alpha_2a)A_2&=0,\\ J_0(\alpha_1b)A_1+Y_0(\alpha_1b)B_1&=0,\\ J_1(\alpha_1a)A_1+Y_1(\alpha_1a)B_1-J_1(\alpha_2a)A_2&=0.\end{aligned}\right\} \tag{19}$$

A necessary and sufficient condition that these equations be consistent is that the determinant of the coefficients vanish:

$$\begin{vmatrix} \frac{\alpha_1}{\epsilon_1}J_0(\alpha_1a) & \frac{\alpha_1}{\epsilon_1}Y_0(\alpha_1a) & -\frac{\alpha_2}{\epsilon_2}J_0(\alpha_2a) \\ J_0(\alpha_1b) & Y_0(\alpha_1b) & 0 \\ J_1(\alpha_1a) & Y_1(\alpha_1a) & -J_1(\alpha_2a) \end{vmatrix}=0. \tag{20}$$

If $\beta$ is specified in advance, then $\alpha_1$ and $\alpha_2$ are known from (14), and (20) specifies a necessary and sufficient relationship between $a$ and $b$. The manner of using this information will be illustrated by a later example.

With (20) satisfied, any two of the three relations given in (19) may be used to solve for the ratios of the constants $A_1$, $B_1$, and $A_2$. By using the first and the third,

$$A_1=K_AA_2, \quad B_1=K_BA_2, \tag{21}$$

where

$$\left.\begin{aligned} K_A&=-\frac{\pi\epsilon_1a}{2}\begin{vmatrix} \frac{\alpha_2}{\epsilon_2}J_0(\alpha_2a), & \frac{\alpha_1}{\epsilon_1}Y_0(\alpha_1a) \\ J_1(\alpha_2a), & Y_1(\alpha_1a) \end{vmatrix}\\ K_B&=-\frac{\pi\epsilon_1a}{2}\begin{vmatrix} \frac{\alpha_1}{\epsilon_1}J_0(\alpha_1a), & \frac{\alpha_2}{\epsilon_2}J_0(\alpha_2a) \\ J_1(\alpha_1a), & J_1(\alpha_2a) \end{vmatrix}\end{aligned}\right\} \tag{22}$$

and we have made use of the standard identity

$$Y_0(s)J_1(s)-Y_1(s)J_0(s)=2/\pi s.$$

---

[2] Page 224 of reference 1.

The field components are then given as follows:

Medium 1: $a \leqq r \leqq b$.

$$\left.\begin{aligned} E_{z1} &= A_2 \frac{\alpha_1}{j\omega\epsilon_1}[K_A J_0(\alpha_1 r) + K_B Y_0(\alpha_1 r)]e^{-j\beta z}, \\ E_{r1} &= A_2 \frac{\beta}{\omega\epsilon_1}[K_A J_1(\alpha_1 r) + K_B Y_1(\alpha_1 r)]e^{-j\beta z}, \\ H_{\phi 1} &= A_2[K_A J_1(\alpha_1 r) + K_B Y_1(\alpha_1 r)]e^{-j\beta z}, \end{aligned}\right\}. \tag{23a}$$

Medium 2: $0 \leqq r \leqq a$.

$$\left.\begin{aligned} E_{z2} &= A_2 \frac{\alpha_2}{j\omega\epsilon_2} J_0(\alpha_2 r)e^{-j\beta z}, \\ E_{r2} &= A_2 \frac{\beta}{\omega\epsilon_2} J_1(\alpha_2 r)e^{-j\beta z}, \\ H_{\phi 2} &= A_2 J_1(\alpha_2 r)e^{-j\beta z}. \end{aligned}\right\} \tag{23b}$$

In (23), the constant $A_2$ remains to be determined.

### 1.3 Axial Field as a Function of Transmitted Power

The constant $A_2$ may be evaluated in terms of the peak transmitted power $P_M$ or twice the average power $P$. For this purpose, the complex Poynting vector is integrated over a convenient cross section of the guide. This reduces to

$$P_M = \int_0^b \int_0^{2\pi} E_r H_\phi^* \cdot r d\phi dr,$$

where $H_\phi^*$ is the complex conjugate of $H_\phi$. By (22), $K_A$ and $K_B$ both contain the factor $\alpha_2$ and are therefore pure imaginaries. If $E_r$ and $H_\phi$ are specified to have zero phase angle at $z=0$, then $A_2$ is also a pure imaginary, and $j$ appears only in the exponential factor. The integration then runs

$$P_M = 2\pi A_2^2 \frac{\beta}{\omega\epsilon_1} F(a,b;\epsilon_1,\epsilon_2),$$

where

$$F = \frac{\epsilon_1}{\epsilon_2}\int_0^a r J_1^2(\alpha_2 r)dr + \int_a^b r[K_A J_1(\alpha_1 r) + K_B Y_1(\alpha_1 r)]^2 dr \tag{24}$$

or

$$A_2 = ((P_M/2\pi)(\omega\epsilon_1/\beta F))^{\frac{1}{2}}. \tag{25}$$

At the center of the tube, the maximum value of the axial field $E_{z2}$ is

$$E_{z0} = \left| A_2 \frac{\alpha_2}{\omega\epsilon_2} \right| = \left| \frac{\alpha_2}{\omega\epsilon_2}\left(\frac{P_M}{2\pi}\frac{\omega\epsilon_1}{\beta F}\right)^{\frac{1}{2}} \right|. \tag{26}$$

### 1.4 Further Discussion of the Solution

The determinant (20) may be considered to yield a solution for $b$ when $\alpha_1$ and $\alpha_2$ are known. After defining a set of functions

$$Q_1(s) = \frac{J_0(s)}{Y_0(s)}, \quad Q_2(s) = -j\frac{J_1(js)}{J_0(js)}, \quad Q_3(s) = \frac{J_0(s)}{J_1(s)}, \quad Q_4(s) = \frac{Y_1(s)}{J_1(s)}, \tag{27}$$

we can reduce (20) to the form

$$Q_1(\alpha_1 b) = \frac{1 - j\frac{\alpha_1}{\alpha_2}\frac{\epsilon_2}{\epsilon_1}Q_2(\alpha_2 a)Q_3(\alpha_1 a)}{Q_4(\alpha_1 a) - j\frac{\alpha_1}{\alpha_2}\frac{\epsilon_2}{\epsilon_1}Q_2(\alpha_2 a)\frac{Q_3(\alpha_1 a)}{Q_1(\alpha_1 a)}}. \tag{28}$$

All of these $Q$ functions, with the exception of $Q_2(s)$ have been plotted.[3] $Q_2(s)$ is plotted in Fig. 2 of this paper. The use of these curves simplifies the calculation of $b$.

The chief remaining problem is an evaluation of the function $F(a, b; \epsilon_1, \epsilon_2)$ in (24). The function is integrable in known Bessel functions. By reference to a standard treatise on Bessel functions,[4] it can be shown that

$$F = \frac{1}{\alpha_2^2}F_1(\alpha_2 a) + \frac{K_A^2}{\alpha_1^2}[F_1(\alpha_1 b) - F_1(\alpha_1 a)] + \frac{2K_AK_B}{\alpha_1^2}[F_3(\alpha_1 b) - F_3(\alpha_1 a)] + \frac{K_B^2}{\alpha_1^2}[F_2(\alpha_1 b) - F_2(\alpha_1 a)], \tag{29}$$

where

$$\left.\begin{aligned} F_1(s) &= \frac{s^2}{2}\{[J_1(s)]^2 - J_0(s)J_2(s)\}, \\ F_2(s) - F_2(u) &= \frac{s^2}{2}\{[Y_1(s)]^2 - Y_0(s)Y_2(s)\} - \frac{u^2}{2}\{[Y_1(u)]^2 - Y_0(u)Y_2(u)\}, \\ F_3(s) &= \tfrac{1}{2}\{s^2[Y_1(s)J_1(s) + Y_2(s)J_2(s)] - s[Y_1(s)J_2(s) + Y_2(s)J_1(s)]\}. \end{aligned}\right\} \tag{30}$$

Future work may indicate whether or not these results can be simplified.

## 2. TYPICAL EXAMPLE

Suppose a $TM_{0,1}$ wave is to be set up at $\omega = 3\cdot 10^{10}$ per second traveling at one-tenth the velocity of light in vacuum, so that $V_p = 3\cdot 10^7$ meters per second, whence $\beta = (10)^3$ per meter. Assume the dielectric material is such that

$\mu = 4\pi 10^{-7}$ henry per meter,
$\epsilon_1 = (10^{-7}/18\pi)$ farad per meter
(relative dielectric constant $= 200$),
$\epsilon_2 = (10^{-9}/36\pi)$ farad per meter.

Then $\alpha_1 = 10^3$, $\alpha_2 = j10^3$ by (14). Next, let us try using a radius $a = 2$ millimeters $= 2\cdot 10^{-3}$ meter. Then

$$\begin{aligned} \alpha_1 a &= 2, \\ \alpha_2 a &= j2, \\ \alpha_1/\epsilon_1 &= 5.55\cdot 10^{11}, \\ \alpha_2/\epsilon_2 &= j1.11\cdot 10^{14}. \end{aligned}$$

From Fig. 2, $Q_2(j2) = 0.70$, and from reference 1 and (27)

$$\begin{aligned} Q_1(2) &= 0.45, \\ Q_3(2) &= 0.39, \\ Q_4(2) &= -0.19. \end{aligned}$$

Then by (28), $Q_1(\alpha_1 b) = -5.2$ whence

$$\alpha_1 b = 0.73,\ 3.78,\ 6.89,\ \cdots$$

and

$$b = 0.7,\ 3.8,\ 6.9,\ \cdots,\ \text{millimeters}.$$

We take $b = 3.8$ millimeters as the smallest available value of $b > a$. A more careful calculation using tables rather than curves shows that $b = 3.6$ millimeters.

Next, compute $K_A$ and $K_B$ using (22).

$$\begin{aligned} K_A &= j153.3, \\ K_B &= j812.5. \end{aligned}$$

The expressions for the field components (23)

---

[3] Pages 200–203 of reference 1.

[4] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Macmillan Company, New York, 1944), second edition.

then become

$$E_{z1}=A_2 2842[J_0(\alpha_1 r)+5.30\,Y_0(\alpha_1 r)]e^{-j\beta z},$$
$$E_{r1}=jA_2 2892[J_1(\alpha_1 r)+5.30\,Y_1(\alpha_1 r)]e^{-j\beta z},$$
$$H_{\phi 1}=A_2 153.3[J_1(\alpha_1 r)+5.30\,Y_1(\alpha_1 r)]e^{-j\beta z},$$
$$E_{z2}=A_2 3703 J_0(\alpha_2 r)e^{-j\beta z},$$
$$E_{r2}=A_2 3774 J_1(\alpha_2 r)e^{-j\beta z},$$
$$H_{\phi 2}=A_2 J_1(\alpha_2 r)e^{-j\beta z}.$$

These quantities are plotted as functions of $r$ in Figs. 3 and 4. Here it is plain that the electric vectors have the same orders of magnitude in both dielectrics, while the magnetic vector is practically negligible in the central dielectric. This fact, combined with the small cross-sectional area of the inner dielectric, results in most of the transmitted power being propagated in the outer dielectric.

The factor $F$ in (24) may be evaluated by numerical integration, whence it turns out that $F=-0.00037-0.3475=-0.3479$, the smaller term being the contribution of the inner dielectric. Since by (26) the transmitted power is related linearly to $F$, it can be seen that only about 0.1 percent of the power is propagated in the central region.

By substitution in (26) we now have immediately $E_{z0}=5.77(P_M)^{\frac{1}{2}}$, volts per centimeter, at the axis of the guide.

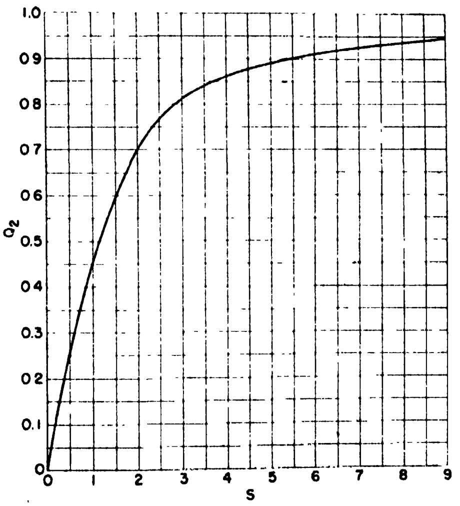


FIG. 2. A useful ratio of certain Bessel functions of imaginary argument. $Q_2=-j[J_1(js)/J_0(js)]$.

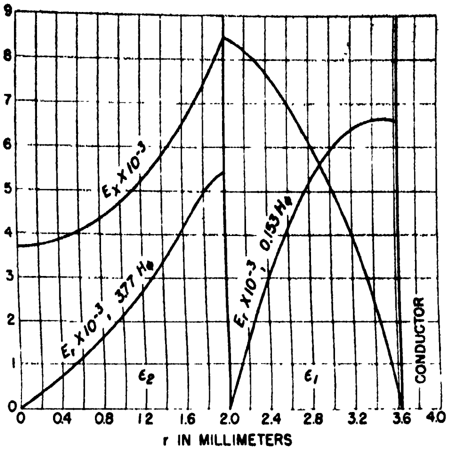


FIG. 3. Distribution of field components in a typical example.

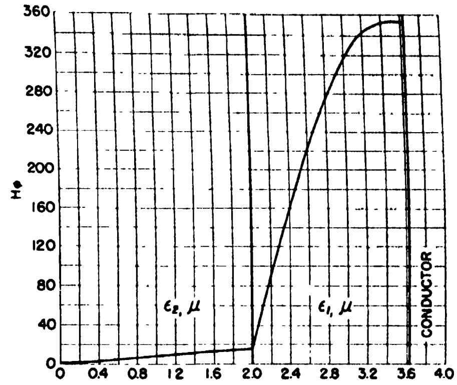


FIG. 4. Distribution of magnetic intensity in a typical example.

# Graphical Methods for Evaluating Fourier Integrals

W. J. Cunningham
*Cruft Laboratory, Harvard University, Cambridge, Massachusetts**

(Received March 12, 1947)

The Fourier transform method of determining the response of a linear system to an arbitrary input signal often has its practical usefulness impaired because of difficulties in evaluating the necessary integrals. One possibility of overcoming these difficulties lies in the application of graphical methods to the transformations. Three such graphical procedures are described, all based upon fundamental properties of the transforms. Each method involves an analysis of the curves of the function to be transformed as a sum of curves of simpler functions whose transforms are known. The methods are useful in cases where the problem is too complicated for a simple analytic solution, or where part of the necessary data is available only in the form of a curve obtained, say, from experimental measurements of transmission characteristics or wave shapes. The accuracy of the methods is restricted only by that of the graphical plotting and curve fitting. If only approximate results are required, they may be obtained relatively quickly by these methods.

## I. INTRODUCTION

One of the standard mathematical methods for determining the response of a linear transmission system to an input disturbance of arbitrary shape is based upon the Fourier integral.[1,2] In the application of this method it is necessary first to express the input function of time as a corresponding function of frequency by means of the appropriate Fourier integral. The resulting frequency function, or transform, is multiplied by the steady-state transmission characteristics of the system to give the frequency function at the output of the system. The time function, or inverse transform, at the output is obtained from the output frequency function by means of the inverse form of the Fourier integral. The integrals needed in these transformations may be written in a number of ways, a convenient form being as follows:

$$g(f)=\int_{-\infty}^{\infty} e(t)\exp(-j\omega t)dt, \qquad (1)$$

$$e(t)=\int_{-\infty}^{\infty} g(f)\exp(+j\omega t)df, \qquad (2)$$

$$\omega=2\pi f,\quad j=+(-1)^{\frac{1}{2}},$$

where $g(f)$ and $e(t)$ are corresponding transforms of frequency and of time. In Fourier integrals of this form, the entire frequency and time scales, both positive and negative values, must be considered.

Several difficulties may arise when an effort is made to apply this method to the solution of practical problems. Often the input signal is known in the form of a curve, but a mathematical expression for it is not available. The same thing may be true (perhaps more often so) for the transmission characteristics of the system. Even if these mathematical expressions were known, it might not be possible to carry out the integrations needed in the evaluation of the Fourier transforms. A number of standard forms of these integrals are available,[3] but if the unknown system does not fit one of these transforms, solution may be very difficult.

## II. GRAPHICAL METHODS

A possible way of obviating these difficulties is to resort to a graphical analysis which will replace the integrations normally required. Furthermore, such graphical methods allow the use of data known only as a curve found from experimental measurements. If extremely accurate results are not required, approximations may be introduced in such a way that the graphical method may yield a useful result more quickly than other rigorous methods. The graphical method is particularly useful when such

---

* Now at Department of Electrical Engineering, Yale University, New Haven, Connecticut.

[1] E. A. Guillemin, *Communication Networks* (John Wiley and Sons, 1935), Vol. II, Chap. XI.

[2] W. L. Sullivan, "Analysis of systems with known transmission frequency characteristics by Fourier integrals," Elec. Eng. 61, 248 (1942).

[3] G. A. Campbell and R. M. Foster, "Fourier integrals for practical applications," Bell Telephone Monograph, B-584, (1931).

approximations are being made, since it is relatively easy to visualize their effect upon the final results. When the system is complicated, the graphical solution likewise becomes complicated, as does any other means of transient analysis. At least two graphical methods for determining the response of a system to an arbitrary input signal have already been described.[4, 5] However, the procedures of the present discussion differ from these two and may have certain advantages in their application.

These methods are based upon fundamental properties of the Fourier integrals.[3] Perhaps the most important property is the superposition principle for linear systems. This principle states that if a function is made up of any linear combination of other functions, the Fourier transform of the first function is equal to the same linear combination of the Fourier transforms of the other functions. Therefore, it is possible to collect as geometric curves a number of mathematically simple functions and their transforms. A function in the form of a complicated curve may be broken into some linear combination of these simple curves. The transform of the complicated curve is then the same linear combination of the transforms of the simple curves. The whole process may be carried out by graphical methods.

In physical problems the input function of time and the output function of time are generally both real. Their corresponding transforms are usually complex functions, each having both a real and an imaginary part. The real parts of these complex functions have even symmetry; their inverse transforms are also real and have even symmetry. The imaginary parts of the complex functions have odd symmetry; their inverse transforms are real and have odd symmetry. Thus, the inverse transform of a complex function is real, and consists of the sum of an even function and an odd function.

An important type of problem is that in which the input to the system is zero until zero time, when the input function begins to have a finite value. If the physically realizable system is passive, its output must also be zero for times less than zero. This fact, plus the discussion of the preceding paragraph, leads to an important conclusion. The real time function at the output of the system must be zero for negative times and will generally have a non-zero value for positive times. Therefore, the even and odd real functions which compose it must have identical geometric shape at positive times and be mirror images of one another at negative times.[6] If this property is utilized, it is often necessary to consider only one of these component functions, and not both simultaneously.

The solution of a typical problem might proceed as shown in Fig. 1, where both positive and negative scales of time and frequency are shown. The input function of time, Fig. 1(a), is applied at zero time and is assumed to be sufficiently simple mathematically so that its transform, Fig. 1(b), may be found as a complex function of frequency from Eq. (1). If this is not true, the problem is more involved, but still may be solved by the methods to be described. It will also be assumed that the steady-state transmission characteristics of the system, Fig. 1(c), are known as a complex function of frequency. These characteristics may be found by the usual steady-state analysis methods, or in the case of actual systems, may be determined experimentally. It is necessary now to find the product, Fig. 1(d), at each frequency, of the complex functions of frequency representing the

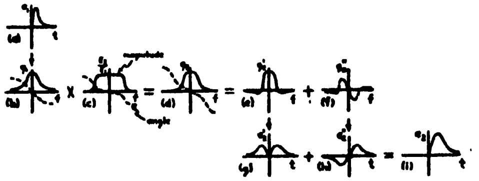


FIG. 1. Outline of solution of typical problem by Fourier integral method. Necessary mathematical operations are shown, with the arrow indicating a (direct or inverse) Fourier transformation. (a) Input signal to system (real). (b) Fourier transform of (a) (complex). (c) Steady-state transmission characteristics of system (complex). (d) Product of (b) and (c) (complex). (e) Real part of (d) (even symmetry). (f) Imaginary part of (d) (odd symmetry). (g) Inverse transform of (e) (real and even). (h) Inverse transform of (f) (real and odd). (i) Output signal from system (real).

---

[4] A. V. Bedford and G. L. Fredendall, "Transient response of multistage video frequency amplifiers," Proc. I.R.E. 27, 277 (1939).

[5] H. A. Wheeler, "The interpretation of amplitude and phase distortion in terms of paired echoes," Proc. I.R.E. 27, 384 (1939).

[6] E. C. Cherry, "Pulse response, a new approach to a.c. electric network theory and measurement," J. Inst. Elec. Eng. 92, III, 183 (1945).

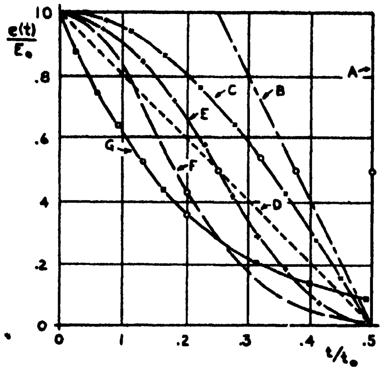


FIG. 2. Real even functions of time; only positive half of each is shown. The maximum value of the function is $E_0$; its total duration is approximately $t_0$. Circles indicate the nominal duration, $t_n$. Corresponding transforms are shown in Fig. 3. A. Rectangle. B. Trapezoid. C. Half cosine. D. Triangle. E. Cosine squared. F. Gaussian. G. Exponential.

input signal and the transmission characteristics of the system. If these two functions are available in algebraic form (magnitude and angle are most convenient), they may be multiplied directly. If one or both of them is available only as a curve, the multiplication of ordinates must be done on a point by point basis. The product obtained in this way will generally be a complex function of frequency and should be expressed as a real part, Fig. 1(e), and an imaginary part, Fig. 1(f). The inverse transform of each part, Fig. 1(g) and (h), is the contribution of that part to the real function of time at the output of the system. As previously pointed out, each contribution must be the same for positive values of time. Because of their even and odd symmetry, their sum at negative times is zero. In order to obtain the output signal, Fig. 1(i), the two contributions may be added, or more simply, only one contribution (either (g) or (h)) need be considered at positive times only, and its value doubled.

The last step of transforming the function of frequency at the output of the system into the corresponding function of time is usually the step producing the greatest mathematical difficulty. It is for this purpose that the methods described here will be found useful.

## III. CORRESPONDING FUNCTIONS

A number of corresponding simple functions of time and of frequency are given in Fig. 2 and Fig. 3. A collection of curves of this sort is useful, either in the direct graphical solution of more complicated wave shapes, or in providing the investigator with a qualitative picture of the relation between typical functions. All of these functions of time and frequency are real and, therefore, must have even symmetry. Only one half of each curve is plotted, for the sake of convenience, but it must be remembered that each of these curves is symmetrical about the vertical axis and has the same values for negative as for positive abscissas. The curves are obtained by assuming relatively simple mathematical functions of time and finding the corresponding functions of frequency with the aid of Eq. (1). For convenience, the curves are plotted with dimensionless ratios as coordinates, and the time functions (Fig. 2) are adjusted so that the approximate total duration of each is the time $t_0$. Also indicated on the curves, with small circles, are the points corresponding to the nominal duration, $t_n$, and the nominal cut-off frequency, $f_n$, for each curve.[5] The nominal duration and nominal cut-off may be defined for real even functions in the following way,

$$\left.\begin{aligned} t_n &= 1/e(0)\int_0^\infty e(t)dt, \\ f_n &= 1/g(0)\int_0^\infty g(f)df, \end{aligned}\right\} \qquad (3)$$

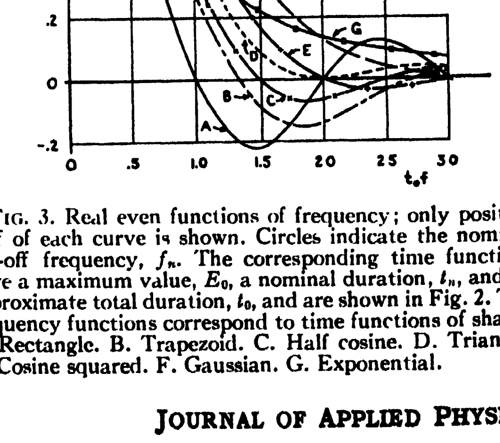


FIG. 3. Real even functions of frequency; only positive half of each curve is shown. Circles indicate the nominal cut-off frequency, $f_n$. The corresponding time functions have a maximum value, $E_0$, a nominal duration, $t_n$, and an approximate total duration, $t_0$, and are shown in Fig. 2. The frequency functions correspond to time functions of shape: A. Rectangle. B. Trapezoid. C. Half cosine. D. Triangle. E. Cosine squared. F. Gaussian. G. Exponential.

where $e(0)$ is $e(t)$ evaluated at zero time, and $g(0)$ is $g(f)$ evaluated at zero frequency, and both $e(0)$ and $g(0)$ are finite. In geometrical terms, the nominal duration, $t_n$, is the base of a rectangle which has an altitude, $e(0)$, and an area equal to the total area under the curve of $e(t)$, taken for positive times only. A similar relation applies to $f_n$. For zero values of the arguments, the quantities are

$$e(0)=\int_{-\infty}^{\infty} g(f)df=2f_n g(0),$$

$$g(0)=\int_{-\infty}^{\infty} e(t)dt=2t_n e(0),$$

and their product gives the fundamental relation,

$$4f_n t_n=1.$$

If this relation is observed and the parameters, $E_0$, $t_0$, and $t_n$, are chosen properly, the scales of Figs. 2 and 3 may be made to fit any desired conditions or systems of units. The maximum value of the time function, occurring here at zero time, is $E_0$.

One further important property of the transforms may be found from inspection of the Fourier integrals. If the arguments of the time function and the frequency function are interchanged, the functions will still correspond provided the sign of one argument is changed. With real even functions, such as Fig. 1(e) and (g), this sign change produces no visible change in the functions. Therefore, the real even functions of Fig. 2 may be taken as functions of frequency, if desired, whereupon the real even functions of Fig. 3 become functions of time. It is sufficient merely to interchange the $t$ and $f$ symbols of these curves.

Another group of corresponding functions which is of interest is shown in Fig. 4. Here, the functions of time, Fig. 4(a) and (c), are cosine functions in which an integral number, $n$, of quarter cycles occurs within the arbitrary time, $t_a$. Since the cosines are even functions the total duration of the functions of time is $2t_a$. Again, the frequency functions, Fig. 4(b) and (d), corresponding to the assumed time functions, may be found with the help of Eq. (1). The form of the result is slightly different depending upon whether $n$ is chosen odd or even. Two families of curves are plotted in Fig. 4 for this reason. These frequency functions are the frequency spectra of a.c. pulses containing a total of $n$ half-cycles, and occurring within the time $2t_a$. As the value of $n$ is made larger, the apparent frequency of the time function is increased, and the main hump of the corresponding frequency spectrum is moved to a point higher on the frequency scale. All of these functions are real and even, although only the positive half of each curve is shown, and once more it is possible to interchange time and frequency if desired.

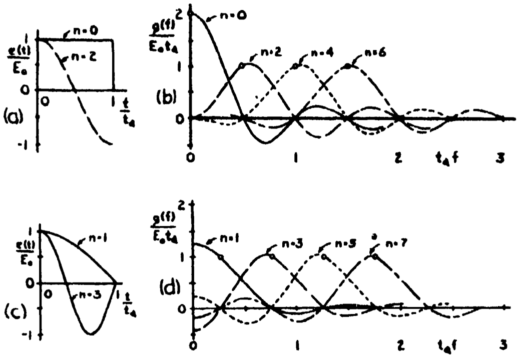


FIG. 4. Corresponding real even functions of time and frequency; only positive half of each curve is shown. Time functions have $n$ quarter cycles in the interval, $t_a$. The parameter, $n$, is even at (a) and odd at (c). Corresponding frequency functions are shown at (b) and (d), respectively. Circles indicate points where only a single curve is not zero.

An important feature of these frequency functions, which is used to simplify later analyses, becomes apparent after a study of Fig. 4. If $n$ is chosen either odd or even, there is a certain frequency at which the frequency function for a single value of $n$ is not zero, while the functions for all other values of $n$ are zero. The function which is not zero at this particular frequency has a value which is a constant, regardless of $n$. In mathematical terms, if $n$ is an integer, either odd or even as the case may be, and $m$ is an integer odd or even to match $n$, then

$$\left.\begin{aligned}&\text{at } f=m/(4t_a), && g(f)=0, && \text{if } m\neq n\\ &\text{at } f=n/(4t_a), && g(f)=E_0t_a, && \text{if } n\neq 0\\ &\text{at } f=0, && g(f)=2E_0t_a, && \text{if } n=0.\end{aligned}\right\}\quad(4)$$

The points where $m=n$ are shown by circles on Fig. 4(b) and (d). In one of the uses of these functions it is necessary to sum a number of

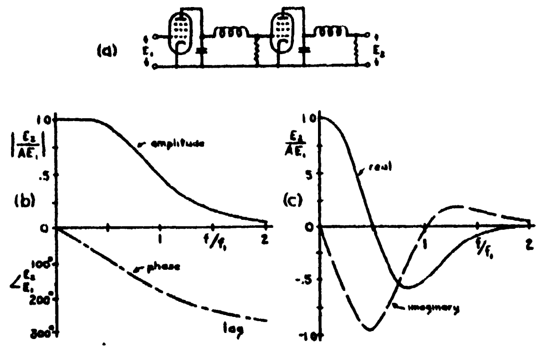


FIG. 5. (a) Low-pass amplifier circuit for use in examples. $E_1$ and $E_2$ are r.m.s. voltages of a single frequency. $A$ is the zero-frequency voltage amplification. $f_1$ is a reference frequency. (b) Steady-state transmission characteristics of (a), amplitude and phase. (c) Steady-state transmission characteristics of (a), real and imaginary parts.

curves having different values of $n$. These points, where $m=n$, are uniquely determined by a single curve and, therefore, may be easily located.

## IV. EXAMPLE FOR TYPICAL SOLUTIONS

As an example to illustrate the application of the graphical methods, the circuit of Fig. 5(a) will be used. This circuit consists of two identical sections of a low-pass network, coupled by pentode vacuum tubes. The r.m.s. voltage of a given frequency applied at the input of the system is $E_1$, and the resulting r.m.s. voltage at the output is $E_2$. For a particular choice of the component element values, the steady-state transmission characteristics of the system are shown in Fig. 5(b) and (c). Figure 5(b) gives these characteristics in terms of amplitude and phase, while Fig. 5(c) represents the same characteristics expressed as a real and an imaginary part. These parts are the products of the amplitude by the cosine and sine, respectively, of the phase angle. At a given frequency, the real part of the transmission characteristics is the ratio between the r.m.s. value of that component of the output voltage, which is in phase with the input voltage, and the r.m.s. value of the input voltage. Similarly, the imaginary part of the transmission characteristics is the corresponding ratio involving that component of the output voltage which leads the input voltage by ninety degrees. In the figure only positive frequencies are shown, although the shapes of the curves for negative frequencies must be understood. The amplitude curve and the real curve have even symmetry; the angle curve and the imaginary curve have odd symmetry. The scales of these curves are given as dimensionless ratios. The factor, $A$, is the voltage amplification of the system at zero frequency, and the reference frequency, $f_1$, is the frequency at which the inductor and capacitor of the circuits are in series resonance.

A non-sinusoidal input voltage, which has an instantaneous value, $e_1(t)$, is now applied to the system. In order not to complicate the example too much, this voltage is assumed to be an impulse function, having an infinite amplitude, zero duration, but a finite area equal to $E_0t_0$. The transform, $g_1(f)$, of such an impulse is easily shown to be a constant, $E_0t_0$, at all frequencies. This fact might be obtained from the curves of Fig. 2 and Fig. 3 for the rectangular wave shape, by allowing the duration, $t_0$, to approach zero as the amplitude, $E_0$, becomes infinite, like $E_0/t_0$. The product of this constant value multiplied by the steady-state transmission characteristics of the system therefore has the shape of these characteristics, with the ordinates merely changed by the constant factor, $E_0t_0$. This product is the frequency function, $g_2(f)$, at the output of the system. Its real part appears in Figs. 6(a), 7(a), and 8(a).

## V. DIRECT CURVE-FITTING METHOD

Perhaps the most direct graphical method of obtaining the output signal, $e_2(t)$, is to make use of the curves of Figs. 2 and 3, combining them graphically so as to fit the frequency function in

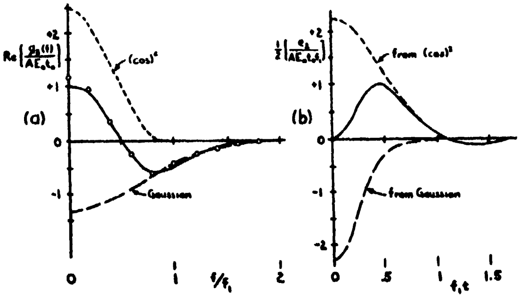


FIG. 6. Solution by direct curve fitting. (a) Real part of frequency function at output of system, as sum of simple curves. Circles indicate sum of dotted curves. (b) Output signal from system, as sum of inverse transforms of simple curves.

question. This procedure is illustrated in Fig. 6. The solid curve in Fig. 6(a) is the real part of the frequency function at the output of the system of Fig. 5(a), when an impulse of area $E_0t_0$ is applied at the input. Again, dimensionless ratios are used on the coordinate axes, and only the positive half of each curve is shown. This curve may be approximated by the sum of the cosine-squared curve and the (negative) Gaussian curve of Fig. 2, with their scales suitably adjusted. Although the curves of Fig. 2 are given as functions of time, it has already been pointed out that their arguments may be changed to frequency if desired. It is convenient to make the change for the present purpose. This curve-fitting is a trial and error process, but is not too difficult. The sum of the two curves is indicated by the circles in Fig. 6(a) and is seen to fit the given curve fairly well.

In Fig. 6(b) the transforms of the two standard frequency functions are plotted, having been obtained with the help of Fig. 3, and with careful observance of the coordinates. Again, only the positive half of each curve is shown. The solid curve of Fig. 6(b) is the sum of the two standard transforms. It has the proper shape for the output signal from the system, but the ordinates have just half the correct value. If the ordinates are doubled, this curve represents the desired output signal.

Where the shape of the frequency function is not too complicated, as is true here, it may be fitted reasonably well by a simple combination of two or three of the standard curves. After some practice this may be done with a minimum of trial and error. Obviously, the more closely the sum of the standard curves approximates the unknown curve, the more accurate will be the final results.

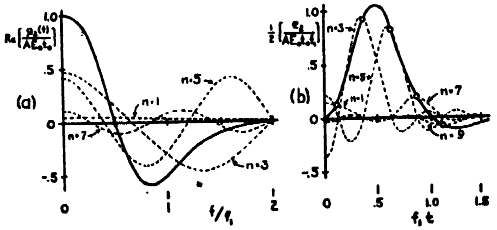


FIG. 7. Solution by cosine analysis of frequency function. (a) Real part of frequency function at output of system, as sum of cosines. (b) Output signal from system, as sum of transforms of cosines.

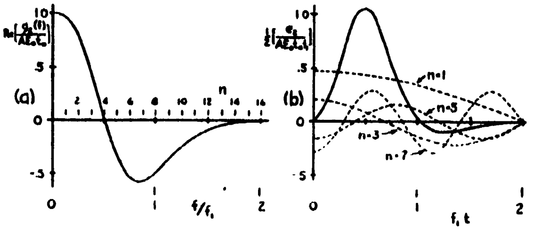


FIG. 8. Solution by cosine synthesis of time function. (a) Real part of frequency function at output of system, with selected points indicated. (b) Output signal from system, as sum of cosines.

## VI. COSINE ANALYSIS OF THE FREQUENCY FUNCTION

In order to eliminate the trial and error of the type of curve-fitting just described, the cosine functions of Fig. 4 and their transforms may be used. It is necessary to interchange time and frequency in these curves, but since they are real even functions, the interchange may be made with no other modifications. The frequency function at the output of the system may be analyzed as a sum of cosine functions by any of a number of standard graphical methods. For example, the Fischer-Hinnen method[7] is well suited for this purpose. The transforms of the cosine functions may then be found from Fig. 4 and summed.

This process is illustrated in Fig. 7. The solid curve in Fig. 7(a) represents the real part of the frequency function at the output of the system, just as before, and is plotted in the same way. This curve is analyzed as the sum of cosines having odd numbers of quarter cycles in the frequency interval, $f_a = 2f_1$. This choice of this interval must be such that the given curve has essentially zero value for frequencies outside of the interval. The cosine functions thus obtained are plotted in Fig. 7(a) with the parameter, $n$, indicating the number of quarter cycles executed in the interval. Only values of $n$ up to seven are plotted, although higher values are actually required to represent the curve well.

The transforms of each of these cosine functions may be found with the help of Fig. 4. The transforms are plotted in Fig. 7(b), showing only values of $n$ up to nine, and are summed. Once

---

[7] F. A. Laws, *Electrical Measurements* (McGraw-Hill Book Company, Inc., New York, 1938), second edition, pp. 687–695.

again, if the ordinates of this sum are doubled, the output signal from the system is obtained. The larger the number of cosines used in the original analysis, the more accurately they will approximate the frequency function, and the more accurate will be the final result.

This analysis of Fig. 7 was carried out with $n$ chosen odd, and the curves of Fig. 4(c) and (d) were used. Exactly the same results will be obtained if $n$ is chosen even, although the analysis is different and the curves of Fig. 4(a) and (b) would be employed. There is usually little basis for choosing between these two analyses since either will produce the same result with about the same amount of effort.

The shape of the transforms in Fig. 4(b) and (d) is relatively complicated, and their plotting and summing as in Fig. 7(b) is tedious. However, if it is sufficient to obtain only a few discrete points on the curve of the output signal, the properties of these functions are such that their plotting may be eliminated. It was pointed out earlier, in connection with Eq. (4), that there are certain times at which only the curve for a single value of $n$ contributes to the final sum. These points are shown by circles in Fig. 7(b). The times at which these points occur are $n/(4f_a)$, and the values at these times are $a_n f_a$, where $a_n$ is the amplitude of the $n$th cosine function of the analysis, and $n$ is not zero. If $n$ is zero, the value at $t=0$ is $2a_0 f_a$. These points are located by data from the cosine analysis, and plotting of the transforms is not required. Evidently, if it is desired to obtain a large number of points closely spaced along the curve of the output function, the frequency interval, $f_a$, must be chosen relatively large and the cosine analysis carried to large values of $n$.

## VII. COSINE SYNTHESIS OF TIME FUNCTION

A further simplification in the analysis results if the foregoing procedure is applied in reverse order. The unknown frequency function is analyzed as the sum of functions having the shape of those of Fig. 4(b) and (d), and their transforms then found as cosine functions. In this case, the frequency and time scales are left as in Fig. 4. This type of analysis is simpler than the analysis into cosines, since it involves only a direct reading of the value of the curve at certain points. It is not necessary to plot the curves into which the frequency function is analyzed.

This process is illustrated in Fig. 8, where once more Fig. 8(a) represents the real part of the frequency function at the output of the system. The abscissa of this curve is divided into frequency intervals, $f_b$, chosen in this case as $f_1/8$. These intervals are numbered consecutively as shown, with $n$ equal to zero at zero frequency. The value of the curve at points corresponding to the various values of $n$ may then be read directly. The transform of the curve is the sum of the ordinates of a number of cosine curves which execute $n$ quarter cycles in the time interval, $t_b = 1/(4f_b) = 2/f_1$, where $n$ is chosen either odd or even. The amplitude of the $n$th cosine is $p_n/t_b$ where $p_n$ is the value of the frequency function at $nf_b$, and $n$ is not zero. If $n$ is zero, the amplitude of the zeroth cosine (i.e., constant term) is $p_0/(2t_b)$. Cosine time functions, with $n$ odd and values only up to seven, are plotted in Fig. 8(b) and summed. Their sum, with its ordinates doubled, is the output signal from the system. The same result would have been obtained had $n$ been chosen even instead of odd, and once more there is little basis for choice in the matter. If the final time function is to be relatively accurate, the frequency interval, $f_b$, must be chosen small enough so that information is obtained from each important part of the frequency function.

## VIII. INPUT SIGNAL IN THE FORM OF A STEP FUNCTION

In the preceding examples, the input signal to the system has been chosen as an impulse function, since this function has the simplest possible transform. For many applications, in the analysis of the behavior of servomechanisms[8] for instance, the step function is often used for test purposes. The step function, $e_s(t)$, is defined as

$$\begin{cases} e_s(t)=0 & \text{for} \quad t<0 \\ e_s(t)=E_0 & \text{for} \quad t>0. \end{cases}$$

The step function may also be written $e_s(t)$

---

[8] H. Lauer, R. Lesnick, L. E. Matson, *Servomechanism Fundamentals* (McGraw-Hill Book Company, Inc., New York, 1947), Chap. IV.

$=E_0/2+e_s''(t)$ where $e_s''(t)$ is defined as

$$\begin{cases} e_s''(t) = -E_0/2 & \text{for} \quad t<0 \\ e_s''(t) = +E_0/2 & \text{for} \quad t>0 \end{cases}$$

and is a real odd function of time. The usual step function, $e_s(t)$, is thus replaced by a constant, $E_0/2$, and the symmetrical step, $e_s''(t)$. The transform of $e_s''(t)$ may be found from Eq. (1) as $g_s''(f) = -jE_0/\omega$, and is an imaginary odd function of frequency. This result is obtained by multiplying the time function by factors of the form $\exp(\pm ct)$, where $c$ is a real constant, the sign of which is chosen so as to make the product vanish at plus and minus infinity. After completing the integration of Eq. (1), $c$ is allowed to become zero, whereupon the transform given above is found.

The response of a system to the symmetrical step function, $e_s''(t)$, may be determined by any of the methods already described in connection with the impulse function. The response to the constant, $E_0/2$, is simply the product of $A$ (the amplification of the system at zero frequency) and $E_0/2$. Finally, the sum of the responses to the symmetrical step, $e_s''(t)$, and to the constant, $E_0/2$, is the response of the system to the step function, $e_s(t)$. If only the real part of the frequency function at the output is used in the solution, the ordinates of the result must be doubled. For a passive system, at rest at $t=0$, the response must also be zero at $t=0$. After a long time the response must be $AE_0$.

## IX. FINAL REMARKS

In each of the three graphical methods described here only the real part of the frequency function at the output of the system was utilized. It would be possible, of course, to make the analysis in terms of the imaginary part alone. There are occasions when both the real and the imaginary parts of the function must be considered. This is true, for example, if a real time function having no symmetry is to be transformed into its corresponding frequency function. In such a case, the frequency function will be complex, and both real and imaginary parts must be found. The necessary transformations may follow any of the procedures already described, since the same methods allow transformation from frequency to time, or *vice versa*.

The imaginary part may be handled most easily by a procedure analogous to the third method described, in which the real part was treated in terms of a sum of cosine functions. Inasmuch as the imaginary part must have odd symmetry, it must be treated in terms of a sum of sine functions. It is not difficult to set up families of sine functions and their transforms analogous to the cosine functions of Fig. 4. A negative sine curve has as a transform a curve similar in shape to those of Fig. 4(b) and (d), with the main hump positive. The curves necessarily have odd symmetry, and one function is imaginary while the other is real. The most important fact so far as an analysis in sine functions is concerned, is that the relations expressed in Eq. (4) for the transforms of cosine functions are also valid for the transforms of sine functions. Therefore, the analysis in terms of sines proceeds in the same way as has been discussed here in terms of cosines. The only differences are the opposite type of symmetry and the negative algebraic sign which must appear in such a ratio as $-p_n/t_b$, giving the amplitude of the $n$th sine for synthesis by the third method.

Only the very simple cases of an impulse function and a step function at the input of the system were considered here. If the input signal is more complicated so that its transform is not readily calculable, a graphical analysis of the type already described may be used to find this transform. If the input signal is real, but has no symmetry, steps (d) through (i) of Fig. 1 must be carried out, in reverse order, of course, to yield the frequency function, $g_1(f)$, corresponding to the input signal. With this function known, the remainder of the analysis proceeds as already described.

Of the three graphical methods discussed here, the third one in which the output signal is synthesized as a sum of cosine (or sine) functions is probably the most useful. However, if the curves are relatively simple, the first method, involving an experimental fitting of the curves, may lead more quickly to a result. As in all graphical methods, the accuracy of the results obtained from any of the procedures depends largely upon the skill with which the plotting is carried out. If only the approximate shape of the output signal is required, the graphical curve-fitting may be simplified so that it may be done in a short time. If accurate results are needed then more careful analysis is required.

Finally, it should be pointed out that the third method is in many problems essentially the same as that used by Bedford and Fredendall.[4] However, the bases for the two methods are somewhat different, and only in certain cases do the two become identical. In the present method, the frequency function at the output of the system is considered in terms of its real or imaginary parts, leading to the synthesis of the output signal as a sum of cosines or sines, all of which have zero relative phase displacement. In the method used by Bedford and Fredendall, the output signal is synthesized as a sum of cosines which have phase displacements determined by the problem. For a given number of components, a better representation may be obtained by the latter method, since an additional variable is allowed. On the other hand, the manipulation required to obtain the result may be simpler in the method described here.

### X. ACKNOWLEDGMENT

The original work reported here was done under the able guidance of Professor E. L. Chaffee and his helpful comments and suggestions are gratefully acknowledged.

---

# Two Calculating Machines for X-Ray Crystal Structure Analysis

A. D. Booth
*Institute for Advanced Study, Princeton, New Jersey, and Birkbeck College, England*

(Received July 2, 1946)

The calculation of structure factors in x-ray structure analysis is one of the most laborious operations. Two mechanisms are described which have proved of great service in several analyses. The simpler of the devices is of general application and can be constructed with comparatively limited workshop facilities.

## INTRODUCTION

APART from large and expensive machines for performing automatically the whole range of crystallographic calculation, there remains a need for simple *ad hoc* devices to deal with particular aspects of the problem. If the current schemes for centralizing the latter stages of Fourier refinement come to fruition, this demand for the "home made" and simple type of calculator is likely to be increased.

It is the purpose of this paper to describe two such arrangements which the author designed and found useful in practical structure analysis. The first is especially valuable in the case of tetragonal space groups and found extensive application during the determination of the structure of pentaerythritol tetranitrate.[1] The second is considerably simpler in principle and in design and has been in constant and satisfactory use in structure analysis[2,3] for the past four years.

## THE TETRAGONAL CALCULATOR

In tetragonal space groups it is necessary to obtain the value of expressions of the type:

$$F(hkl)=\sum_r f_r\left[\cos 2\pi\frac{hx_r}{a}\cdot\cos 2\pi\frac{ky_r}{b}-\cos 2\pi\frac{kx_r}{a}\cdot\cos 2\pi\frac{hy_r}{b}\right]\cos 2\pi\frac{lz_r}{c} \quad (1)$$

for a range of values of $(h, k, l)$. The information given being the values of $(hx_r/a,\ ky_r/b,\ lz_r/c)$ and $f_r$.

This calculation is performed by the mechanism shown in Fig. 1.

$L_1N_1$ and $L_2N_2$ are two bars of $\frac{3}{8}''$ diam. steel

---

[1] Booth and Llewellyn, Proc. Roy. Soc. (to be published).

[2] G. A. Jeffrey, Proc. Roy. Soc. A183, 388 (1945).

[3] G. A. Jeffrey, Proc. Roy. Soc. A188, 222 (1947).

rod which are capable of rotation in a horizontal plane about vertical axes through their centers, $O_1$ and $O_2$, respectively. Sliding along these bars are riders $P_1$ and $P_2$, so arranged as to be capable of fixation at any given radii $OP_1$ and $OP_2$ (positive or negative). Attached to the riders is a link mechanism $VP_1RP_2$; this has the bars $P_1RP_2$, $VR$ rigidly fixed at right angles, and the links $VP_1$, $VP_2$ of equal length. The bar $P_1RP_2$ can slide freely through the rider points $P_1P_2$, whilst the links $P_1V$, $P_2V$ can rotate about $P_1$, $P_2$, and $V$ with the latter freely sliding on $RV$.

The effect of this mechanism is seen to be the maintenance of $R$ at the mid point of $P_1$ and $P_2$.

If now the bars $L_1N_1$, $L_2N_2$ are rotated through angles $\theta_1$ and $\theta_2$ from the horizontal line $O_1O_2$, as shown, the displacement of $R$, parallel to $O_1O_2$, is simply:

$$\tfrac{1}{2}(O_2P_2\cos\theta_2 - O_1P_1\cos\theta_1). \tag{2}$$

This component of the motion of $R$ is separated by means of the framework $ABCD$ which can move, on rollers, in a horizontal line parallel to $O_1O_2$, $R$ moving in a groove in $AB$.

The motion of $CD$ is communicated, *via* a silk string $S$, to a disk $E$ which rotates about a vertical axis through its center $M$. The top surface of $E$ is covered with fine calico and has applied to it a planimeter assembly $P$. If the axis of the planimeter makes an angle $\theta_3$ with the radius $Mp$ through its point of contact $p$, the resultant rotation is

$$\Theta\cdot Mp\cdot\cos\theta_3, \tag{3}$$

where $\Theta$ is the angle of rotation of $E$.

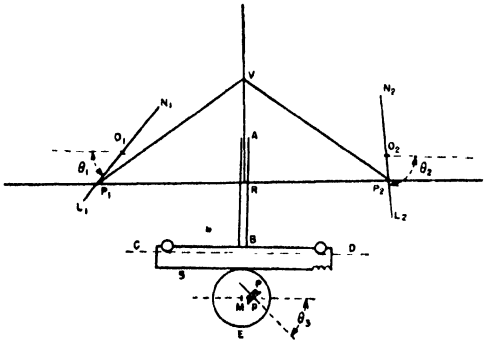


FIG. 1.

Thus, if scale factors are properly adjusted, the motion recorded on the planimeter is:

$$Mp(O_2P_2\cos\theta_2 - O_1P_1\cos\theta_1)\cos\theta_3. \tag{4}$$

The planimeter assembly is provided with a clutch so that it can be removed from contact with $E$.

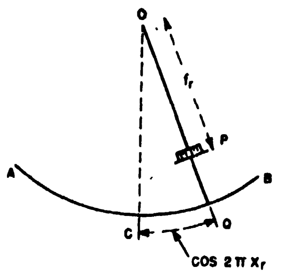


FIG. 2.

It follows, that by making:

$$Mp = f_r$$

$$O_2P_2 = \cos 2\pi\frac{hx_r}{a},$$

$$\theta_2 = \cos 2\pi\frac{ky_r}{b},$$

$$O_1P_1 = \cos 2\pi\frac{kx_r}{a},$$

$$\theta_1 = \cos 2\pi\frac{hy_r}{b},$$

$$\theta_3 = 2\pi\frac{lz_r}{c},$$

and resetting between operations, the summation (1) can be computed. The bars $L_1N_1$, $L_2N_2$ bear upon their upper surfaces scales directly calibrated in $\cos 2\pi(hx/a)$.

The machine proved most effective when two operators were available, one for each turntable, $L_1N_1$, $L_2N_2$, and reduced the time of computing the complicated expression (1) by a factor of approximately 4.

Unfortunately no photographs of the machine were taken and it has now, with the ending of the war, been dismantled.

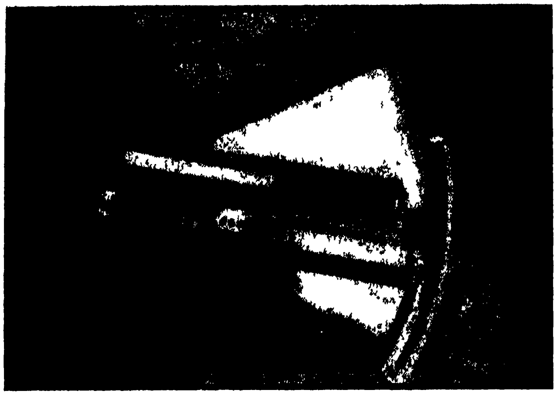


FIG. 3.

### THE COSINE ADDING MACHINE

This is considerably simpler in principle than the machine just described and has the advantage of not requiring any delicate mechanism of the planimeter type.

Although not always the most mathematically elegant method, it is easily shown that any structure factor formula can be reduced to a set of summations of the type:

$$\sum_r f_r \begin{Bmatrix} \sin \\ \cos \end{Bmatrix} 2\pi\left(\frac{hx_r}{a} \pm \frac{ky_r}{b} \pm \frac{lz_r}{c}\right). \quad (5)$$

A table of the functions $hx_r/a \pm ky_r/b + lz_r/c$ is readily computed using a "Facit" or other multiplying machine so that it is desirable to have a simple and rapid means of effecting the summation:

$$\sum_r f_r \begin{Bmatrix} \sin \\ \cos \end{Bmatrix} 2\pi X_r. \quad (6)$$

This is found in the mechanism shown in Fig. 2. A planimeter wheel $P$ is mounted on a bar $OQ$ pivoted at $O$: if the bar is turned through an angle $CQ = K\cos 2\pi X_r$, where $K$ is constant, the motion recorded by $P$ is simply:

$$K' \cdot OP \cdot \cos 2\pi X_r.$$

The constants are so arranged that $OP \propto f_r$ so that:

$$f_r \cos 2\pi X_r$$

is recorded.

By removing $P$ from contact with the plane $OABC$ during the return of $Q$ to its zero, $C$, and resetting $OP$, the complete summation (6) can be rapidly effected.

The machine is particularly advantageous in organic structure analysis where $f_r$ is constant for a large number of atoms.

Practical details are shown in Fig. 3. The planimeter wheel has a vulcanized rubber tire, the outside diameter being 1.625″. The counter mechanism was extracted from an old electricity meter with a step up ratio between wheel and counter of 1:7. The angle scale of $\cos 2\pi X_r$ is in the form of a circular segment and is attached to the planimeter on rail assemblies as shown. A simple spring push type clutch is provided so that the planimeter is normally out of contact with the friction plate, the latter being of sand blasted brass. An audible "click" is given on the return of the setting to zero.

This machine proves not much inferior in speed to a normal key driven adding machine.

### DISCUSSION

The second, and simpler of these machines, has been and is of constant utility and is sufficiently versatile to be worth making a permanent fixture in a crystallographic laboratory. The first is of more ephemeral application, but is worthy of description as containing a number of mechanical elements which can be incorporated in other special purpose machines, and which may consequently be of service to other workers in the field.

The author wishes to record his thanks to Mr. H. W. Giles for his care and numerous helpful suggestions while constructing the second machine, and to the Board of the British Rubber Producers' Research Association under whose auspices the machine was constructed. For the first machine valuable mechanical help was given by Mr. W. H. Small, B.Sc.

# The Field of a Microwave Dipole Antenna in the Vicinity of the Horizon*

C. L. Pekeris**
*Columbia University Mathematical Physics Group, New York, New York*

(Received April 8, 1947)

In this paper a method is developed for determining the electromagnetic field produced by a microwave antenna at points on the horizon, and on either side of it, where neither the ray theory nor the normal mode theory can be used conveniently. The theory is developed for a condition of standard atmospheric refraction, by use of a space in which the earth is flattened and the rays are curved. This allows us to make a simple derivation of the ray theory, valid in the optical region, and of the normal mode theory, suited for the shadow zone. For the intermediate region centered around the horizon we use the original integral for the potential to obtain expressions for the field under the restriction of maximum absorption, which for typical ground conditions applies to wave-lengths less than about a meter. Three cases are treated in which the transmitter, or receiver, are either situated on the ground or are elevated several natural units of height. For an elevated transmitter and receiver the Hertzian potential $\Psi$ due to a point source at the origin is in the vicinity of the horizon given by

$$|\Psi| = [1/2(r\hat{r})^{\frac{1}{2}}]\left|(2/\pi^{\frac{1}{2}})e^{i\pi/4}\int_0^{\infty} \exp[-i(2\tau^{\frac{1}{2}}t+t^2)]dt - [(2/3)^{\frac{1}{2}}\Omega/\pi^{\frac{1}{2}}]e^{-i\pi/12}F(p)\right|, \quad \text{(A)}$$

with

$$p=(3/2)^{\frac{2}{3}}(x-\mathfrak{x}),\ \Omega^2=(z_1^{-\frac{1}{2}}+z_2^{-\frac{1}{2}}),$$
$$\tau=(x-\mathfrak{x})^2/\Omega^2,\ \mathfrak{x}=z_1^{\frac{1}{2}}+z_2^{\frac{1}{2}},$$

where $x$ denotes the horizontal distance $r$ expressed in natural units, $z_1$ and $z_2$ the heights of transmitter and receiver in natural units, $\hat{r}$ the distance of receiver from transmitter when the former is on the horizon. $F(p)$ (see Eq. (68) below) has been evaluated, and is given in Table IV, while the integral in (A) can be expressed in terms of the tabulated Fresnel integrals. In the limit of very short wave-lengths the field *on* the horizon approaches the value $1/(2\hat{r})$ which would result from the diffraction of the *direct ray only* by a straight edge placed at the point of tangency of the horizon with the earth. A comparison of the field obtained from (A) with exact values computed by van d. Pol and Bremmer, using the ray theory and the normal mode theory, is shown in Figs. 6 and 7.

When the transmitter is at zero elevation and the receiver is elevated several units of height, the potential in the vicinity of the horizon is given by

$$|\Psi_v| = \frac{(3/2\pi a_e)^{\frac{1}{3}}\lambda^{\frac{1}{3}}}{\pi(r\hat{r})^{\frac{1}{2}}}\left|\frac{G(p)\epsilon_1}{(\epsilon_1-1)^{\frac{1}{2}}}\right|, \quad \text{(B)}$$

$$|\Psi_h| = \frac{(3/2\pi a_e)^{\frac{1}{3}}\lambda^{\frac{1}{3}}}{\pi(r\hat{r})^{\frac{1}{2}}}\left|\frac{G(p)}{(\epsilon_1-1)^{\frac{1}{2}}}\right| \quad \text{(C)}$$

for vertical polarization and horizontal polarization, respectively. Here $\epsilon_1$ denotes the complex dielectric constant, $a_e$ the effective radius of the earth, and $\lambda$ the wave-length. $G(p)$ is given in Eq. (78) and is shown in Fig. 4. A comparison of (B) with exact values obtained by van d. Pol and Bremmer is shown in Fig. 8.

When both the transmitter and receiver are at zero elevation, it is found that the potential can be expressed as the *sum* of the surface wave appropriate for a *flat* ground and an integral depending on the radius of the earth. At great distances, the two terms tend to cancel out. Under conditions of maximum absorption this leads to

$$|\Psi_v| = \frac{2}{r^2k_0}\left|\frac{\epsilon_1^2 g(p')}{(\epsilon_1-1)}\right|; \quad |\Psi_h| = \frac{2}{r^2k_0}\left|\frac{g(p')}{(\epsilon_1-1)}\right|, \quad \text{(D)}$$

$$p'=(3/2)^{\frac{2}{3}}x, \quad g(p')=1-(2e^{i\pi/4}/3\pi^{\frac{1}{2}})(p')^{\frac{3}{2}}H(p').$$

$H(p')$ is given in (84) and $g(p')$ is shown in Fig. 5, where it is compared with results obtained previously by van d. Pol and Bremmer using the normal mode theory.

For points *on* the horizon Eq. (A) reduces to

$$|\Psi| = (1/2\hat{r})|1-(0.684-0.183i)\Omega|, \quad \text{(E)}$$

while in (B) and (C) we put $|G(0)| = 2.13$.

## 1. INTRODUCTION

In the treatment of the problem of diffraction of radio waves around the earth[1] there are available at present the ray method, which is valid only in the lit region, and the method of normal modes, which is suited for the shadow region beyond the horizon. On the horizon and beyond the ray method breaks down completely, while the method of normal modes requires an increasing number of modes to be used as one recedes from the horizon into the lit

---

* This paper is based on work done for the Naval Research Laboratory under contract with the Navy's Office of Research and Inventions. Publication assisted by the Ernest Kempton Adams Fund for Physical Research of Columbia University.

** On leave of absence from Massachusetts Institute of Technology.

[1] G. N. Watson, Proc. Roy. Soc. **A95**, 83 (1918). B. van d. Pol and H. Bremmer, Phil. Mag. **24**, 141, 825 (1937); **25**, 817 (1938); **27**, 261 (1939). B. Wwedensky, Tech. Phys. U.S.S.R. **2**, 624 (1935); **3**, 915 (1936); **4**, 579 (1937). T. L. Eckersley and G. Millington, Phil. Trans. Roy. Soc. **237**, 273 (1938). P. S. Epstein, Proc. Nat. Acad. Sci. **21**, 62 (1935). G. Millington, Phil. Mag. **27**, 517 (1939). C. R. Burrows and M. C. Gray, Proc. I.R.E. **29**, 16, 1941. K. A. Norton, F. C. C. Report 39920 (1940); F. C. C. Report 47475 (1941).

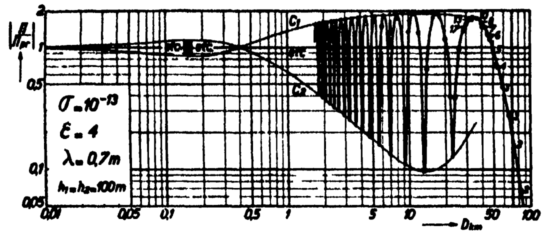


FIG. 1. Black circle points were calculated by the ray theory. The numbers placed near the open circles represent the number of modes that had to be included in order to obtain an accuracy of one percent. The horizon is at the dashed line (van d. Pol and Bremmer, Phil. Mag. 27, 270 (1939)).

region. This is illustrated in Fig. 1 which is taken from the work of van d. Pol and Bremmer.[2] The purpose of this investigation is to develop a theory of diffraction of microwaves which is intrinsically most accurate in the vicinity of the horizon, and which becomes less accurate as one recedes from the horizon either far into the lit zone or far into the shadow zone, thus bridging the gap between the ray theory and the normal mode theory. The horizon is the locus of coalescence of the direct ray with the reflected ray, and for this reason our work displays some points of similarity with Airy's theory of diffraction near a caustic. Expression (A) for the Hertzian potential given above reduces to an interference pattern between the direct and reflected rays for points inside the lit zone (Figs. 6 and 7), while inside the shadow zone it is a monotone decreasing function caused by cancellation of the two terms.

The propagation of microwaves around the earth presents a peculiar diffraction problem, in that the distance of the source from the sphere, as well as the wave-length, is small in comparison with the radius of the sphere. Furthermore, we are interested in the diffracted field at points which are within a range of only a fraction of the radius from the sphere. The question has often been raised to what extent an object of finite curvature diffracts like a straight edge placed at the point of tangency of the horizon line. In this connection our analysis shows that, within ranges of the order of a fraction of the radius from the sphere, the field *on the horizon* approaches the value for a straight edge in the limit of short wave-lengths, provided we neglect the contribution from the reflected ray. The latter is justified because of the divergence of the reflected ray near the horizon.

Our treatment differs from Watson's, and its later extensions by van d. Pol and Bremmer,[1] in that the problem of diffraction around the spherical earth is transformed into one of diffraction over a flat earth through an atmosphere with a suitably variable index of refraction. This device was first introduced by Schelleng, Burrows and Ferrell,[3] and it is known to be valid[4] out to ranges of the order of the radius of the earth and for wave-lengths greater than a centimeter up to heights of the order of a thousand feet. By this method one obviates the synthesis of an original solution expressed as a series of spherical harmonics, and, furthermore, arrives directly at an integral for the potential containing Bessel functions of the order $\frac{1}{3}$, which are intrinsic to the problem.

The effect of the curvature of the earth is introduced by replacing the actual index of refraction $\mu$ by a modified index $N$ defined by

$$N(r) = r\mu(r)/a\mu(a), \qquad (1)$$

$a$ denoting the radius of the earth and $r$ the distance from the center of the earth. In a free atmosphere $N(r)$ increases with height $h$ above the surface at the rate of $h/a$, while under conditions of standard atmospheric refraction this rate is changed to $h/a_e$, with[3] $a_e \simeq 4a/3$. Our problem reduces therefore to finding a solution for the Hertzian potential $\Psi$ satisfying the wave equation:

$$\nabla^2\Psi + k_0{}^2\mu^2\Psi = 0, \quad k_0 = 2\pi/\lambda, \qquad (2)$$

$$\mu^2 = (1+qh), \quad q = 2/a_e, \quad \text{in air}, \qquad (3)$$

$$\mu^2 = (1+qh)\epsilon_1, \quad \epsilon_1 = \epsilon - i\sigma/\omega, \quad \text{in the ground}, \qquad (4)$$

where $\epsilon$ denotes the dielectric constant and $\sigma$ the conductivity of the ground, and a time factor $e^{i\omega t}$ has been assumed. It will be noted that it is necessary to distort the space inside the earth as well. The fact that the expression for $\mu^2$ in (4) becomes formally negative inside the earth (where $h$ is negative) at a depth of about half $a_e$

---

[2] B. van d. Pol and H. Bremmer, Phil. Mag. 27, 270, (1939).

[3] J. C. Schelleng, C. R. Burrows, and E. B. Ferrell, Proc. I.R.E. 21, 427 (1933); J. E. Freehafer, Radiation Laboratory Report 447 (1943).

[4] C. L. Pekeris, Phys. Rev. 70, 518 (1946).

need not concern us, since the earth-flattening approximation is to be applied only under conditions where the energy inside the earth is confined to a thin surface layer.

It should be noted that the problem treated here finds an application in underwater acoustics,[5] where the index of refraction $\mu$ for sound propagation sometimes varies with depth $h$ according to Eq. (3). In that problem of a real flat boundary and a variable $\mu$, the parameter $q$ in (3), which is a measure of the curvature of the rays, need not be small. On the other hand, in the application to propagation of microwaves around the earth, the earth-flattening approximation is valid only insofar as powers of $qh$ higher than the first can be neglected, so that in the sequel, the expression for the field should be relied on only as far as the first power in $qh$.

## 2. FORMAL SOLUTION

The solution of (2) must reduce to the form $e^{-ikR}/R$ at small distances $R$ from the source, it must represent an outgoing wave at great distances from the source, and the tangential components of the electric and magnetic fields must be continuous at the surface of the earth. The formal solution of (2), satisfying the above mentioned boundary conditions, can be obtained by a method given by H. Lamb.[5,6] The solution is built up from elementary solutions of the form

$$\psi = e^{i\omega t} J_0(kr) F(h) G(k), \tag{5}$$

$r$ now denoting horizontal distance, and $k$ a variable over which one eventually integrates in the complex $k$-plane. $F(h)$ satisfies the equations

$$(d^2F/dh^2) + [k_0^2(1+qh) - k^2]F = 0, \quad \text{in the air } (h>0), \tag{6}$$

$$(d^2F/dh^2) + [k_1^2(1+qh) - k^2]F = 0; \quad k_1^2 = k_0^2\epsilon_1, \quad \text{in the ground } (h<0). \tag{7}$$

The fundamental solutions of (6) are

$$M(u) = u^{\frac{1}{2}} H_{1/3}{}^{(1)}(2u^{\frac{3}{2}}/3), \quad N(u) = u^{\frac{1}{2}} H_{1/3}{}^{(2)}(2u^{\frac{3}{2}}/3), \quad u = (k_0/q)^{\frac{2}{3}}[1+qh-(k^2/k_0^2)], \tag{8}$$

while the solution of (7), which does not become infinite at great depth, is

$$P(v) = v^{\frac{1}{2}}[J_{1/3}(2v^{\frac{3}{2}}/3) + J_{-1/3}(2v^{\frac{3}{2}}/3)] = [(3v)^{\frac{1}{2}}/2][H_{1/3}{}^{(1)}(2v^{\frac{3}{2}}/3)e^{i\pi/6} + e^{-i\pi/6}H_{1/3}{}^{(2)}(2v^{\frac{3}{2}}/3)],$$
$$v = (k_1/q)^{\frac{2}{3}}[1+qh-(k^2/k_1^2)]. \tag{9}$$

Here the phase of $u$ is zero when positive, and $-i\pi$ when negative, and similarly for $v$. Let the transmitter's height above the ground be denoted by $h_1$, then we have

$$\begin{aligned} &F_1(h) = C(k)N(u), \quad h>h_1, \\ &F_2(h) = A(k)M(u) + B(k)N(u), \quad 0<h<h_1, \\ &F_3(h) = E(k)P(v), \quad h<0, \end{aligned} \tag{10}$$

where the functions of $k$, $A$, $B$, $C$, and $E$ are determined by the boundary conditions. These are, for a point source at $h_1$,

$$F_1 = F_2, \quad (\partial F_2/\partial h) - (\partial F_1/\partial h) = 2k, \quad \text{at } h = h_1, \tag{11}$$

$$\partial F_2/\partial h = \partial F_3/\partial h, \quad k_0^2 F_2 = k_1^2 F_3 \quad \text{at } h=0, \text{ for vertical polarization}, \tag{12}$$

$$\partial F_2/\partial h = \partial F_3/\partial h, \quad F_2 = F_3 \quad \text{at } h=0, \text{ for horizontal polarization}. \tag{13}$$

Using the relations

$$dM/du = uH_{-2/3}{}^{(1)}(2u^{\frac{3}{2}}/3), \quad dN/du = uH_{-2/3}{}^{(2)}(2u^{\frac{3}{2}}/3), \tag{14}$$

$$H_\nu{}^{(1)}(z) \cdot H_{\nu+1}{}^{(2)}(z) - H_{\nu+1}{}^{(1)}(z) \cdot H_\nu{}^{(2)}(z) = 4i/\pi z, \tag{15}$$

---

[5] C. L. Pekeris, J. Acous. Soc. Am. **18**, 295 (1946).

[6] H. Lamb, Phil. Trans. Roy. Soc. **A203**, 1 (1904); W. H. Furry, Radiation Laboratory Report 680 (1945). See also Arnold Sommerfeld's article in Frank-Mises, *Differential-gleichungen der Physik* (Rosenberg, New York, 1943) Vol. II, p. 918.

in solving for $A$, $B$, $C$, and $E$ from (11) and (12), or from (11) and (13), we obtain

$$\Psi_1=-[i\pi e^{\omega t}/3(qk_0^2)^{1/3}]\int_0^\infty J_0(kr)kdk(u_1u)^{1/2}H_{1/3}^{(2)}(2u^{3/2}/3)[H_{1/3}^{(1)}(2u_1^{3/2}/3)+$$
$$+(B/A)H_{1/3}^{(2)}(2u_1^{3/2}/3)],\quad h>h_1, \tag{16}$$

$$\Psi_2=-[i\pi e^{\omega t}/3(qk_0^2)^{1/3}]\int_0^\infty J_0(kr)kdk(u_1u)^{1/2}H_{1/3}^{(2)}(2u_1^{3/2}/3)[H_{1/3}^{(1)}(2u^{3/2}/3)$$
$$+(B/A)H_{1/3}^{(2)}(2u^{3/2}/3)],\quad 0<h<h_1, \tag{17}$$

$$\Psi_3=-[i\pi\gamma e^{\omega t}/3(qk_0^2)^{1/3}]\int_0^\infty J_0(kr)kdk[P(v)/P(v_0)]H_{1/3}^{(2)}(2u^{3/2}/3)[H_{1/3}^{(1)}(2u_0^{3/2}/3)+$$
$$+(B/A)H_{1/3}^{(2)}(2u_0^{3/2}/3)],\quad h<0, \tag{18}$$

$$u_0=(k_0/q)^{2/3}[1-(k^2/k_0^2)],\quad u_1=(k_0/q)^{2/3}[1+qh_1-(k^2/k_0^2)],\quad v_0=(k_1/q)^{2/3}[1-(k^2/k_1^2)], \tag{19}$$

$$B/A=-[\beta u_0^{1/2}H_{1/3}^{(1)}(2u_0^{3/2}/3)-u_0H_{-2/3}^{(1)}(2u_0^{3/2}/3)]/[\beta u_0^{1/2}H_{1/3}^{(2)}(2u_0^{3/2}/3)-u_0H_{-2/3}^{(2)}(2u_0^{3/2}/3)], \tag{20}$$

$$\beta=(1/\epsilon_1^{2/3})[\dot{P}(v_0)/P(v_0)]\simeq(iv_0^{1/2}/\epsilon_1^{2/3});\quad \gamma=1/\epsilon_1,\ \text{for vertical polarization}, \tag{21}$$

$$\beta=\epsilon_1^{1/3}[\dot{P}(v_0)/P(v_0)]\simeq iv_0^{1/2}\cdot\epsilon_1^{1/3};\quad \gamma=1,\ \text{for horizontal polarization}. \tag{22}$$

The only assumption made in deriving this solution, aside from the one involved in the earth-flattening approximation, was in writing $iv_0^{1/2}$ for $\dot{P}(v_0)/P(v_0)$. This approximation holds with increasing accuracy the more absorptive the ground is; even for dry soil the error introduced by this approximation is less than one part in 10[7].

We shall now show that in the limit of an infinite radius of the sphere ($q\to0$) our solution reduces to Sommerfeld's solution[6] for the potential of a dipole over a *flat* earth. As $q\to0$, $u_0$ defined in (19) becomes very large, and

$$u_1^{3/2}\to u_0^{3/2}+(3h_1/2)(k_0^2-k^2)^{1/2}+\cdots;\quad u\to u_0^{3/2}+(3h/2)(k_0^2-k^2)^{1/2}+\cdots. \tag{23}$$

We can therefore replace the Hankel functions in (16) and (17) by their asymptotic expansions:

$$H_{1/3}^{(1)}(z)\to(2/\pi z)^{1/2}\exp[i(z-5\pi/12)],\quad H_{1/3}^{(2)}(z)\to(2/\pi z)^{1/2}\exp[i(-z+5\pi/12)], \tag{24}$$

$$B/A\to\exp[-i5\pi/6+4iu_0^{3/2}/3]R(k),\quad k<k_0, \tag{25}$$

$$R(k)=\left[\frac{k_1^2(k_0^2-k^2)^{1/2}-k_0^2(k_1^2-k^2)^{1/2}}{k_1^2(k_0^2-k^2)^{1/2}+k_0^2(k_1^2-k^2)^{1/2}}\right],\ \text{for vertical polarization,}$$
$$=\left[\frac{(k_0^2-k^2)^{1/2}-(k_1^2-k^2)^{1/2}}{(k_0^2-k^2)^{1/2}+(k_1^2-k^2)^{1/2}}\right],\ \text{for horizontal polarization,} \tag{26}$$

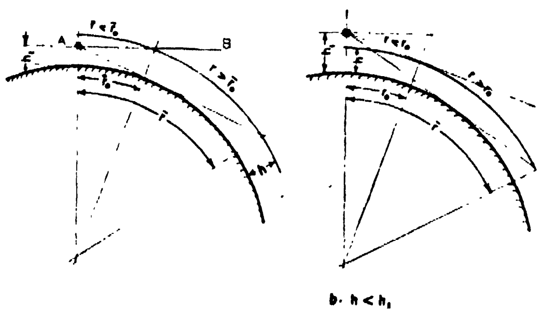
$$(u_1u)^{1/2}H_{1/3}^{(2)}(2u^{3/2}/3)[H_{1/3}^{(1)}(2u_1^{3/2}/3)+(B/A)H_{1/3}^{(2)}(2u_1^{3/2}/3)]$$
$$\to\frac{3(qk_0^2)^{1/3}}{\pi(k_0^2-k^2)^{1/2}}\{\exp[-i(h-h_1)(k_0^2-k^2)^{1/2}]+R(k)\exp[-i(h_1+h)(k_0^2-k^2)^{1/2}]\}, \tag{27}$$

$$\Psi_1\to-i\int_0^\infty\frac{J_0(kr)kdk}{(k_0^2-k^2)^{1/2}}\{\exp[-i(h-h_1)(k_0^2-k^2)^{1/2}]+R(k)\ \exp[-i(h_1+h)(k_0^2-k^2)^{1/2}]\},\quad h>h_1. \tag{28}$$

This agrees with Sommerfeld's solution for a flat earth in which $h_1\to0$, since $(k_0^2-k^2)^{1/2}\to-i(k^2-k_0^2)^{1/2}$ for $k>k_0$.

---

[7] Reference 5, p. 296.



b. $h < h_1$

FIG. 2. $h_1$ = height of transmitter, $h$ = height of receiver.

### 3. RAY THEORY

In this section we shall show briefly how one can derive from our solution given in (16) and (17) the geometric-optical approximation, which is valid for short wave-lengths inside the lit region. No detailed discussion of this subject will be attempted here, since this has been done adequately originally by van d. Pol and Bremmer and later by Burrows and Norton,[1] and since, furthermore, our main interest in this investigation lies in determining the field in the vicinity of the horizon. Now, according to Fermat's principle, a ray is a curve connecting the transmitter and receiver such that along it the travel-time is stationary. Such a curve is characterized by a definite angle $\theta$ which it makes with the vertical at the transmitter, and this value of $\theta$ is associated with a particular value of $k$ ($<k_0$) in the integrals (16) and (17), through the relation[7] $k = k_0\mu(h_1)\sin\theta$. Hence our task is to find the values of $k$ in the path of integration for which the travel-time is stationary. This travel-time is equal to the limit approached by the phase of the wave as the frequency (or $k_0$) is increased indefinitely, because the *beginning* of a pulse starting from quiescence is controlled by the high frequency components. It follows that our procedure is first to obtain asymptotic expressions for the integrands valid for large wave numbers $k_0$, then find the values of $k$ for which the phase is stationary (if such points exist), and finally to evaluate the integrals around the points of stationary phase.

To begin with, the integrals in (16) and (17) can, under certain conditions[8] which are met in our case, be transformed by the use of the relation

$$\int_0^\infty J_0(kr)\psi(k)kdk = \tfrac{1}{2}\int_{-i\infty}^{\infty} H_0^{(2)}(kr)\psi(k)kdk. \quad (29)$$

In the limit of large $r$ (in comparison with $\lambda$) we have

$$H_0^{(2)}(kr) \rightarrow (2/\pi kr)^{\frac{1}{2}} \exp[i\pi/4 - ikr], \quad (30)$$

and (29) reduces to

$$\int_0^\infty J_0(kr)\psi(k)kdk$$
$$\rightarrow [e^{i\pi/4}/(2\pi r)^{\frac{1}{2}}]\int_{-i\infty}^{\infty} e^{-ikr}\psi(k)k^{\frac{1}{2}}dk. \quad (31)$$

Taking now the integral in (16) we write it in the form

$$\Psi_1 \simeq \frac{\pi \exp[i\omega t - i\pi/4]}{3(qk_0^2)^{\frac{1}{3}}(2\pi r)^{\frac{1}{2}}}$$
$$\times \int_{-i\infty}^{\infty} e^{-ikr}[V(k) + W(k)]k^{\frac{1}{2}}dk, \quad h > h_1, \quad (32)$$

$$V(k) \equiv (u_1u)^{\frac{1}{2}}H_{1/3}^{(2)}(2u^{\frac{3}{2}}/3)H_{1/3}^{(1)}(2u_1^{\frac{3}{2}}/3),$$
$$(33)$$
$$W(k) \equiv (u_1u)^{\frac{1}{2}}(B/A)H_{1/3}^{(2)}(2u^{\frac{3}{2}}/3)H_{1/3}^{(2)}(2u_1^{\frac{3}{2}}/3).$$

For large wave numbers $k_0$, and as long as $k$ is less than $k_0(1+qh_1)^{\frac{1}{2}}$ (and therefore also less than $k_0(1+qh)^{\frac{1}{2}}$), both $u_1$ and $u$ are large, so that we obtain from (24)

$$e^{-ikr}V(k) \rightarrow (3/\pi)(uu_1)^{-\frac{1}{4}} \exp[-ip_1(k)],$$
$$p_1(k) = kr + (2u^{\frac{3}{2}}/3) - (2u_1^{\frac{3}{2}}/3). \quad (34)$$

The point $k$ of stationary phase is determined from

$$\partial p_1(k)/\partial k = r - (2x/q)[(1+qh-x^2)^{\frac{1}{2}}$$
$$-(1+qh_1-x^2)^{\frac{1}{2}}] = 0, \quad (35)$$
$$x \equiv (k/k_0) = (1+qh_1)^{\frac{1}{2}}\sin\theta.$$

As $x$ varies from *zero* to its maximum possible value of $(1+qh_1)^{\frac{1}{2}}$, $r$ changes from *zero* to $\bar{r}_0$, where

$$q\bar{r}_0 = 2[1+qh_1]^{\frac{1}{2}} \cdot [q(h-h_1)]^{\frac{1}{2}} \quad (36)$$

defines the ray which starts out horizontally at the source. In real space the rays defined by (35) are the direct rays emitted above the transmitter's horizon.[9]

---

[8] Reference 5, pp. 305 to 307.

[9] Above line $AB$ in Fig. 2.

If the receiver is located above $h_1$ but below the transmitter's horizon ($r > \bar{r}_0$), the phase of the term $e^{-ikr}V(k)$ in (32) does not have a stationary point, so that this term's contribution becomes small relative to the second term's, on account of its highly oscillatory nature. The direct ray reaching the receiver in this case arises from a stationary point $k$ of $e^{-ikr}W(k)$, where $k_0 < k < k_0(1+qh_1)^{\frac{1}{2}}$. This term has also another stationary point at $k < k_0$, which gives the reflected ray for any position of the receiver in the lit zone. When $k_0 < k < k_0(1+qh_1)^{\frac{1}{2}}$ we put

$$w = (2/3)(k_0/q)[(k^2/k_0^2)-1]^{\frac{3}{2}}, \quad (2u_0^{\frac{3}{2}}/3) = e^{-i3\pi/2}w, \tag{37}$$

and obtain from (20)

$$\frac{B}{A} = \left\{ \frac{\beta[I_{-1/3}(w) + e^{-i\pi/3}I_{1/3}(w)] + (3w/2)^{\frac{1}{3}}[I_{2/3}(w) + e^{-i\pi/3}I_{-2/3}(w)]}{\beta[I_{-1/3}(w) + e^{i\pi/3}I_{1/3}(w)] + (3w/2)^{\frac{1}{3}}[I_{2/3}(w) + e^{i\pi/3}I_{-2/3}(w)]} \right\}. \tag{38}$$

For large $w$

$$I_\nu(w) \rightarrow [e^w/(2\pi w)^{\frac{1}{2}}][1 - \{(4\nu^2-1)/8w\} + \cdots], \quad B/A \rightarrow e^{-i\pi/3}, \tag{39}$$

and

$$e^{-ikr}W(k) \rightarrow (3/\pi)(uu_1)^{-\frac{1}{4}} \exp[(i\pi/2) - ip_2(k)], \quad p_2(k) = kr + (2u^{\frac{3}{2}}/3) + (2u_1^{\frac{3}{2}}/3), \tag{40}$$

$$\partial p_2(k)/\partial k = r - (2x/q)[(1+qh_1-x^2)^{\frac{1}{2}} + (1+qh-x^2)^{\frac{1}{2}}] = 0. \tag{41}$$

As $x$ varies from its minimum value of 1 to its maximum value of $(1+qh_1)^{\frac{1}{2}}$, $r$, as given by (41), changes from $\bar{r}$ to $\bar{r}_0$, where $\bar{r}$ denotes the horizontal distance from transmitter when the receiver is on the horizon, and is given by

$$q\bar{r} = 2(qh_1)^{\frac{1}{2}} + 2(qh)^{\frac{1}{2}}. \tag{42}$$

These are the direct rays emitted below the transmitter's horizon, thus complementing the direct rays not included in $p_1(k)$.

For $k < k_0$ we have, on using (25),

$$e^{-ikr}W(k) \rightarrow R(k)(3/\pi)(uu_1)^{-\frac{1}{4}} \exp[-ip_3(k)], \quad p_3(k) = kr - (4u_0^{\frac{3}{2}}/3) + (2u^{\frac{3}{2}}/3) + (2u_1^{\frac{3}{2}}/3), \tag{43}$$

$$\partial p_3(k)/\partial k = r - (2x/q)[(1+qh_1-x^2)^{\frac{1}{2}} + (1+qh-x^2)^{\frac{1}{2}} - 2(1-x^2)^{\frac{1}{2}}] = 0. \tag{44}$$

Here, as $x$ varies from 0 to 1, $r$ changes from *zero* to $\bar{r}$. These are the rays *reflected* at the surface of the earth, $R(k)$ representing the reflection coefficient for a plane earth, with the divergence factor arising from the integration around the stationary point, as is shown in the Appendix.

When the elevation of the receiver is less than that of the transmitter ($h < h_1$) we have to use $\Psi_2$ given in (17). Here the second term is identical with $W(k)$ in (32), and therefore leads to the phases $p_2(k)$ and $p_3(k)$, which are symmetrical in $h_1$ and $h$. In the case of $p_2(k)$

$$k_0 < k < k_0(1+qh)^{\frac{1}{2}}, \quad q\bar{r} > qr > qr_0 = 2[1+qh]^{\frac{1}{2}} \cdot [q(h_1-h)]^{\frac{1}{2}}, \tag{45}$$

where $r_0$ denotes the horizontal distance when the transmitter is on the horizon of the receiver (see

TABLE I. Location of the points $x(=k^*/k_0)$ of stationary phase, and the rays associated with them.

| | | | Limits of $x$ | | Limits of $r$ | | | Sign of $\frac{\partial^2 p[k^*]}{\partial k^2}$ |
|---|---|---|---|---|---|---|---|---|
| | Phase | Eq. | $x_1$ | $x_2$ | $r(x_1)$ | $r(x_2)$ | Nature of ray | |
| $h > h_1$ | $p_1(k)$ | 35 | 0 | $(1+qh_1)^{\frac{1}{2}}$ | 0 | $\bar{r}_0$ | Direct | $-$ |
| | $p_2(k)$ | 41 | 1 | $(1+qh_1)^{\frac{1}{2}}$ | $\bar{r}$ | $\bar{r}_0$ | Direct | $+$ |
| | $p_3(k)$ | 44 | 0 | 1 | 0 | $\bar{r}$ | Reflected | $-$ |
| $h < h_1$ | $p_4(k)$ | 47 | 0 | $(1+qh)^{\frac{1}{2}}$ | 0 | $r_0$ | Direct | $-$ |
| | $p_2(k)$ | 41 | 1 | $(1+qh)^{\frac{1}{2}}$ | $\bar{r}$ | $r_0$ | Direct | $+$ |
| | $p_3(k)$ | 44 | 0 | 1 | 0 | $\bar{r}$ | Reflected | $-$ |

Fig. 2). When the receiver is closer to the transmitter ($r<r_0$), the direct ray arises from a stationary point of the first term in the integrand of (17):

$$e^{-ikr}(u_1u)^{\frac{1}{2}}H_{1/8}{}^{(2)}(2u_1{}^{\frac{3}{2}}/3)H_{1/3}{}^{(1)}(2u^{\frac{3}{2}}/3)\rightarrow(3/\pi)(uu_1)^{-\frac{1}{4}}\exp[-ip_4(k)]. \tag{46}$$

$$p_4(k)\equiv kr+(2u_1{}^{\frac{3}{2}}/3)-(2u^{\frac{3}{2}}/3),\quad \partial p_4(k)/\partial k=r-(2x/q)[(1+qh_1-x^2)^{\frac{1}{2}}-(1+qh-x^2)^{\frac{1}{2}}]=0. \tag{47}$$

As $x$ varies from *zero* to $(1+qh)^{\frac{1}{2}}$, $r$ changes from *zero* to $r_0$, thus complementing the rays arising from the stationary point in the second term. The reflected ray in the case $h<h_1$ arises from the second term, and is given by (43) and (44). The location of the points of stationary phase is summarized in Table I. Figure 2 illustrates the geometry of the problem.

We shall now indicate how one obtains the amplitudes of the rays by integrating around the points of stationary phase. It will suffice to illustrate the procedure in the case of the first term of $\Psi_2$ in (17), which yields the direct ray given in (47). We have first, on using (31) and then (46),

$$\begin{aligned}\Psi_2'\equiv-\frac{i\pi\exp[i\omega t]}{3(qk_0{}^2)^{\frac{1}{3}}}&\int_0^\infty J_0(kr)kdk(u_1u)^{\frac{1}{2}}H_{1/3}{}^{(2)}(2u_1{}^{\frac{3}{2}}/3)H_{1/3}{}^{(1)}(2u^{\frac{3}{2}}/3)\\ &\simeq\frac{\pi\exp[i\omega t-i\pi/4]}{3(qk_0{}^2)^{\frac{1}{3}}(2\pi r)^{\frac{1}{2}}}\int_{-i\infty}^\infty e^{-ikr}(u_1u)^{\frac{1}{2}}H_{1/3}{}^{(2)}(2u_1{}^{\frac{3}{2}}/3)H_{1/3}{}^{(1)}(2u^{\frac{3}{2}}/3)(k)^{\frac{1}{2}}dk\\ &\simeq\frac{\exp[i\omega t-i\pi/4]}{(qk_0{}^2)^{\frac{1}{3}}(2\pi r)^{\frac{1}{2}}}\int_{-i\infty}^\infty(uu_1)^{-\frac{1}{4}}\exp[-ip_4(k)]k^{\frac{1}{2}}dk.\end{aligned} \tag{48}$$

Let the point $k$ at which $p_4(k)$ is stationary be denoted by $k^*$; then we have in the neighborhood of $k^*$

$$p_4(k)=p_4(k^*)+\tfrac{1}{2}[\partial^2p_4(k^*)/\partial k^2]t^2+\cdots,\quad t=k-k^*. \tag{49}$$

Using this expansion in (48) we obtain

$$\begin{aligned}\int_{-i\infty}^\infty(uu_1)^{-\frac{1}{4}}\exp[-ip_4(k)]k^{\frac{1}{2}}dk&\simeq\exp[-ip_4(k^*)][(uu_1)^{-\frac{1}{4}}k^{\frac{1}{2}}]_{k=k^*}\cdot\int_{-\infty}^\infty\exp[-(i\partial^2p_4/2\partial k^2)t^2]dt\\ &=(k^*)^{\frac{1}{2}}(q/k_0)^{\frac{1}{2}}(1+qh_1-x^2)^{-\frac{1}{4}}(1+qh-x^2)^{-\frac{1}{4}}\exp[i\pi/4-ip_4(k^*)]\left|\frac{2\pi}{\partial^2p_4/\partial k^2}\right|^{\frac{1}{2}}\end{aligned} \tag{50}$$

where we have written $x$ for $k^*/k_0$. Hence, the direct ray is given by

$$\Psi_2'\simeq\frac{\exp[i\omega t-ip_4(k^*)]x^{\frac{1}{2}}}{[k_0r|\partial^2p_4/\partial k^2|]^{\frac{1}{2}}(1+qh_1-x^2)^{\frac{1}{4}}(1+qh-x^2)^{\frac{1}{4}}}. \tag{51}$$

This can now be expressed explicitly in terms of $r$, $h_1$, and $h$ by using (47). It follows from the latter that

$$x^2=[r^2/4R^2]\cdot[a+b+(4ab-q^2r^2)^{\frac{1}{2}}],\quad a\equiv1+qh_1,\quad b\equiv1+qh,\quad R\equiv[r^2+(h-h_1)^2]^{\frac{1}{2}}, \tag{52}$$

$$\partial^2p_4/\partial k^2=-[r(4ab-q^2r^2)^{\frac{1}{2}}/2k_0x(a-x^2)^{\frac{1}{2}}(b-x^2)^{\frac{1}{2}}],$$

$$\Psi_2'\simeq\frac{\exp[i\omega t-ip_4(k^*)]x2^{\frac{1}{2}}}{r(4ab-q^2r^2)^{\frac{1}{4}}}=\frac{\exp[i\omega t-ip_4(k^*)][a+b+(4ab-q^2r^2)^{\frac{1}{2}}]^{\frac{1}{2}}}{2^{\frac{1}{2}}R(4ab-q^2r^2)^{\frac{1}{4}}}, \tag{53}$$

$$p_4(k^*)=[k_0/(3\cdot2^{\frac{1}{2}}R)]\cdot[(a+b+qR)^{\frac{1}{2}}+(a+b-qR)^{\frac{1}{2}}]\cdot[q^2r^2+2a^2+2b^2-(a+b)(4ab-q^2r^2)^{\frac{1}{2}}]. \tag{54}$$

The amplitude of the direct ray reduces to $R^{-1}$ if powers of $q$ higher than the first are neglected.

As to the normal mode theory, which is useful in the shadow zone, we shall merely mention here that it can be obtained by evaluating the integral in (32) at the residues of $(B/A)$ defined in (20). For details of this calculation in the case of strong absorption ($\beta\gg1$) the reader is referred to reference 5, pages 307 and 308.

TABLE II.

| | $k<k_0$ | $k>k_0$ |
|---|---|---|
| $u_0$ | $(3v/2)^{\frac{2}{3}}$ | $e^{-i\pi}(3w/2)^{\frac{2}{3}}$ |
| $u$ | $z_2+(3v/2)^{\frac{2}{3}}$ | $z_2-(3w/2)^{\frac{2}{3}}$ |
| $u_1$ | $z_1+(3v/2)^{\frac{2}{3}}$ | $z_1-(3w/2)^{\frac{2}{3}}$ |
| $v_0(19)$ | $[L^2(\epsilon_1-1)/4H^2\epsilon_1^{\frac{1}{2}}]+(3v/\epsilon_1 2)^{\frac{2}{3}}$ | $[L^2(\epsilon_1-1)/4H^2\epsilon_1^{\frac{1}{2}}]-(3w/\epsilon_1 2)^{\frac{2}{3}}$ |
| $(k/k_0)$ | $1-\frac{H^2}{2L^2}(12v)^{\frac{2}{3}}-\frac{H^4}{8L^4}(12v)^{4/3}+\cdots$ | $1+\frac{H^2}{2L^2}(12w)^{\frac{2}{3}}-\frac{H^4}{8L^4}(12w)^{4/3}+\cdots$ |
| $p_2(k)$ | $p_2(k_0)-pv^{\frac{2}{3}}+[\bar{x}/4(z_1z_2)^{\frac{1}{2}}](3v/2)^{4/3}+\cdots$ | $p_2(k_0)+pw^{\frac{2}{3}}+[\bar{x}/4(z_1z_2)^{\frac{1}{2}}](3v/2)^{4/3}+\cdots$ |

## 4. THE FIELD IN THE VICINITY OF THE HORIZON IN CASE THE TRANSMITTER AND RECEIVER ARE ELEVATED SEVERAL NATURAL UNITS OF HEIGHT ABOVE THE GROUND

We now turn to our principal task, which is the determination of the field in the vicinity of the horizon. It will be noted from Table I that as $r\rightarrow\bar{r}$ (the horizon) the stationary point of $p_2(k)$ approaches $k_0$ from above, while the stationary point of $p_3(k)$ approaches the same limit from below. This corresponds to the coalescence of the direct ray with the reflected ray on the horizon. Since both of these stationary points arise from the second terms (containing the $B/A$ factor) in the integrands of (16) and (17), and since the principal contribution to the integrals occurs near the stationary points, it follows that, except for the case when the transmitter or receiver are on the ground, the field in the neighborhood of the horizon can be determined approximately by integrating the second terms in the integrands of (16) and (17) over a small range of $k$ on either side of $k_0$. When the transmitter or receiver, or both, are on the ground, both terms in the integrands of (16) or (17) need to be considered.

In the sequel it will be convenient to define a natural unit of height, $H$, and a natural unit of horizontal distance, $L$, through

$$H=(k_0^2q)^{-\frac{1}{3}},\quad L=2(kq^2)^{-\frac{1}{3}},\quad z_1=h_1/H,$$
$$z_2=h/H,\quad x=r/L,\quad p=(3/2)^{\frac{1}{3}}(x-\bar{x}).\qquad(55)$$

If λ is expressed in centimeters then, for a mean radius of the earth of 6371 km and $a_e=(4/3)a$, we have

$$H(m)=2.208\lambda^{\frac{2}{3}},\quad L(\text{km})=6.124\lambda^{\frac{1}{3}}.\qquad(56)$$

Further, we write

$$(k^2/k_0^2)-1=(3qw/2k_0)^{\frac{2}{3}},\quad k>k_0;\qquad k_0,\qquad(57)$$
$$1-(k^2/k_0^2)=(3vq/2k_0)^{\frac{2}{3}},\quad k<$$

and tabulate the expansions of some relevant quantities in Table II. $(B/A)$ is given in terms of $w$ for $k>k_0$ in (38), while for $k<k_0$ we have

$$\frac{B}{A}=\left\{\frac{\beta[J_{-1/3}(v)-e^{-i\pi/3}J_{1/3}(v)]+(3v/2)^{\frac{1}{3}}[J_{2/3}(v)+e^{-i\pi/3}J_{-2/3}(v)]}{\beta[J_{-1/3}(v)-e^{i\pi/3}J_{1/3}(v)]+(3v/2)^{\frac{1}{3}}[J_{2/3}(v)+e^{i\pi/3}J_{-2/3}(v)]}\right\}.\qquad(58)$$

We shall limit the discussion to the case of strong absorption when $\beta$ is large, so that

$$\frac{B}{A}\longrightarrow-\left[\frac{H_{1/3}^{(1)}(2u_0^{\frac{3}{2}}/3)}{H_{1/3}^{(2)}(2u_0^{\frac{3}{2}}/3)}\right]=\frac{[J_{-1/3}(v)-e^{-i\pi/3}J_{1/3}(v)]}{[J_{-1/3}(v)-e^{i\pi/3}J_{1/3}(v)]}=\frac{[I_{-1/3}(w)+e^{-i\pi/3}I_{1/3}(w)]}{[I_{-1/3}(w)+e^{i\pi/3}I_{1/3}(w)]}.\qquad(59)$$

Now

$$\beta_v=(iv_0^{\frac{1}{2}}/\epsilon_1^{\frac{1}{2}})\sim i[(k_0/q)^{\frac{1}{3}}]\cdot[(\epsilon_1-1)^{\frac{1}{2}}/\epsilon_1];\quad \beta_h=\beta_v\epsilon_1,\qquad(60)$$

where the subscripts $v$ and $h$ signify vertical and horizontal polarization, respectively. Table III shows in four typical cases of ground constants that for wave-lengths less than a meter $|\beta_v|\gg1$, and that this is true at all frequencies for $|\beta_h|$. Relation (59) is strictly true when the ground is a perfect reflector,[8] i.e., when the boundary condition at the ground is $\Psi=0$.

It follows from the preceding discussion that in the vicinity of the horizon

$$\Psi\simeq\frac{i\pi\exp[i\omega t]}{3(qk_0^2)^{1/3}}\int_0^\infty J_0(kr)kdk(u_1u)^{1/2}H_{1/3}^{(2)}(2u^{3/2}/3)H_{1/3}^{(2)}(2u_1^{3/2}/3)\frac{H_{1/3}^{(1)}(2u_0^{3/2}/3)}{H_{1/3}^{(2)}(2u_0^{3/2}/3)}$$

$$\simeq-\frac{\pi\exp[i\omega t-i\pi/4]}{3(qk_0^2)^{1/3}(2\pi r)^{1/2}}\int_{-i\infty}^\infty e^{-ikr}(u_1u)^{1/2}H_{1/3}^{(2)}(2u^{3/2}/3)H_{1/3}^{(2)}(2u_1^{3/2}/3)\frac{H_{1/3}^{(1)}(2u_0^{3/2}/3)}{H_{1/3}^{(2)}(2u_0^{3/2}/3)}k^{1/2}dk \quad (61)$$

$$\simeq\frac{\pi^{1/2}\exp[i\omega t-i\pi/4]}{3(12)^{1/3}(rL)^{1/2}}\left\{\int_0^\infty e^{-ikr}(u_1u)^{1/2}H_{1/3}^{(2)}(2u^{3/2}/3)H_{1/3}^{(2)}(2u_1^{3/2}/3)\left[\frac{I_{-1/3}(w)+e^{-i\pi/3}I_{1/3}(w)}{I_{-1/3}(w)+e^{i\pi/3}I_{1/3}(w)}\right]\frac{dw}{w^{1/3}}\right.$$

$$\left.+\int_0^\infty e^{-ikr}(u_1u)^{1/2}H_{1/3}^{(2)}(2u^{3/2}/3)H_{1/3}^{(2)}(2u_1^{3/2}/3)\left[\frac{J_{-1/3}(v)-e^{-i\pi/3}J_{1/3}(v)}{J_{-1/3}(v)-e^{i\pi/3}J_{1/3}(v)}\right]\frac{dv}{v^{1/3}}\right\}, \quad (62)$$

where $u$, $u_1$, and $k$ take on their respective expressions in terms of $w$ and $v$, as shown in Table II. The only approximation made in going from (61) to (62) was in writing $k_0^{1/2}$ for $k^{1/2}$, which is justified because $H^2/L^2$ $(=1.3\times10^{-7}\lambda^{2/3}{}_{\text{cm}})$ is extremely small in the microwave region.

We shall evaluate the integral in (62) for the case when $z_1$ and $z$ are several units. We can then write

$$e^{-ikr}(u_1u)^{1/2}H_{1/3}^{(2)}(2u^{3/2}/3)H_{1/3}^{(2)}(2u_1^{3/2}/3)\rightarrow(3/\pi)(uu_1)^{-1/4}\exp[i5\pi/6-ip_2(k)], \quad (63)$$

$$\Psi\simeq\frac{\exp[i7\pi/12+i\omega t-ip_2(k_0)]}{(12)^{1/3}(\pi rL)^{1/2}(z_1z_2)^{1/4}}\left\{\int_0^\infty\exp[-ipw^{2/3}-isw^{1/3}]\left(e^{-i\pi/3}+\frac{e^{i\pi/3}[I_{-1/3}(w)-I_{1/3}(w)]}{[I_{-1/3}(w)+e^{i\pi/3}I_{1/3}(w)]}\right)\frac{dw}{w^{1/3}}\right.$$

$$\left.+\int_0^\infty\exp[ipv^{2/3}-isv^{4/3}]\left(\frac{J_{-1/3}(v)-e^{-i\pi/3}J_{1/3}(v)}{J_{-1/3}(v)-e^{i\pi/3}J_{1/3}(v)}\right)\frac{dv}{v^{1/3}}\right\}, \quad (64)$$

$$\int_0^\infty\exp[-ipw^{2/3}-isw^{4/3}](dw/w^{1/3})=[(12)^{1/3}/\Omega]\int_0^\infty\exp[-i(2\tau^{1/2}t+t^2)]dt, \quad (65)$$

$$\Omega^2=[(1/z_1^{1/2})+(1/z_2^{1/2})],\quad \tau=(x-\bar{x})^2/\Omega^2,\quad s=(3/2)^{4/3}\bar{x}/4(z_1z_2)^{1/2}=(3/2)^{1/3}\Omega^2/4;$$

$$\int_0^\infty\exp[ipv^{2/3}]\left[\frac{J_{-1/3}(v)-e^{-i\pi/3}J_{1/3}(v)}{J_{-1/3}(v)-e^{i\pi/3}J_{1/3}(v)}\right]\frac{dv}{v^{1/3}}$$

$$=e^{i\pi/3}\int_0^\infty\exp\{[(-3^{1/2}/2)+(i/2)]px^{2/3}\}\left[\frac{I_{-1/3}(x)-I_{1/3}(x)}{I_{-1/3}(x)+e^{-i\pi/3}I_{1/3}(x)}\right], \quad (66)$$

$$\Psi=\frac{\exp[i\omega t-ip_2(k_0)]}{2(r\bar{r})^{1/2}}\left\{\frac{2e^{i\pi/4}}{\pi^{1/2}}\int_0^\infty\exp[-i(2\tau^{1/2}t+t^2)]dt-\frac{(2/3)^{1/3}\Omega e^{-i\pi/12}}{\pi^{1/2}}F(p)\right\}, \quad (67)$$

TABLE III. Values of $|\beta_v|$ and $|\beta_h|$ for four typical ground constants.

| $\lambda$ | 1 cm | 10 cm | 100 cm | 10 m | 600 m | |
|---|---|---|---|---|---|---|
| $\lvert\beta_v\rvert$ | 601 | 279 | 129 | 49.8 | 1.87 | |
| $\lvert\beta_h\rvert$ | 2400 | 1120 | 523 | 359 | 672 | $\epsilon=4$, $\sigma=10^{-13}$ e.m.u. |
| $\lvert\beta_v\rvert$ | 436 | 202 | 93.9 | 43.3 | 3.40 | Dry soil |
| $\lvert\beta_h\rvert$ | 3920 | 1820 | 845 | 397 | 368 | $\epsilon=9$, $\sigma=3\times10^{-14}$ e.m.u. |
| $\lvert\beta_v\rvert$ | 346 | 161 | 74.5 | 34.3 | 2.64 | Average soil |
| $\lvert\beta_h\rvert$ | 5190 | 2410 | 1120 | 525 | 476 | $\epsilon=15$, $\sigma=5\times10^{-14}$ e.m.u. |
| $\lvert\beta_v\rvert$ | 154 | 69.3 | 17.0 | 2.53 | 0.084 | Sea water |
| $\lvert\beta_h\rvert$ | 12300 | 5920 | 5260 | 7590 | 15020 | $\epsilon=80$, $\sigma=5\times10^{-11}$ e.m.u. |

TABLE IV. Values of $F(p)$ and $G(p)$ defined in (68) and (78), respectively; $R$ and $I$ denote real and imaginary parts.

| $p$ | −4.0 | −3.5 | −3.0 | −2.5 | −2.0 | −1.5 | −1.0 | −0.5 | 0 | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $RF(p)$ | −2.35 | .18 | 1.70 | 2.21 | 2.21 | 2.04 | 1.84 | 1.64 | 1.44 | 1.23 | 1.03 | .84 | .68 | .55 | .45 | .38 | .33 |
| $IF(p)$ | −2.79 | −3.15 | −2.18 | −1.10 | −.36 | .02 | .14 | .10 | 0 | −.11 | −.20 | −.26 | −.27 | −.26 | −.24 | −.21 | −.19 |
| $RG(p)$ | | | | | 8.38 | 5.96 | 3.62 | 2.49 | 1.84 | 1.24 | .67 | .27 | .07 | | | | |
| $IG(p)$ | | | | | −1.63 | 2.89 | 3.24 | 2.22 | 1.06 | .23 | −.16 | −.23 | −.16 | | | | |

$$F(p)=\int_0^\infty \exp[-ipx^{\frac{2}{3}}]\frac{dx}{x^{\frac{1}{3}}}\left[\frac{I_{-1/3}(x)-I_{1/3}(x)}{I_{-1/3}(x)+e^{i\pi/3}I_{1/3}(x)}\right]$$

$$+\int_0^\infty \exp\{[(-3^{\frac{1}{2}}/2)+(i/2)]px^{\frac{2}{3}}\}\frac{dx}{x^{\frac{1}{3}}}\left[\frac{I_{-1/3}(x)-I_{1/3}(x)}{I_{-1/3}(x)+e^{-i\pi/3}I_{1/3}(x)}\right]. \quad (68)$$

In deriving (67) we have neglected the contribution from the $v^{4/3}$ and $w^{4/3}$ terms in the exponents of the *second* and *third* terms in the integrand of (64). This is justified if the principal contribution to $F(p)$ arises from a small range in $x$ near the origin, or if $s\ll 1$. In any case, the error involved can be estimated by using instead of $F(p)$ in (67) the quantity $[F(p)+is(d^2F/dp^2)]$. If $s(d^2F/dp^2)$ is small, it can be added to $F(p)$ to improve the accuracy; if it is not small, (67) cannot be relied upon.

Values of $F(p)$ at intervals of 0.5 obtained from (68) are shown[10] in Table IV. From the second differences, approximate values of $d^2F/dp^2$ were obtained, and are plotted in Fig. 3.

The integral in (67) can be expressed in terms of the tabulated Fresnel[11] integrals, as follows:

$$(2/\pi^{\frac{1}{2}})(e^{i\pi/4})\int_0^\infty \exp[-i(2\tau^{\frac{1}{2}}t+t^2)]dt=e^{i\tau}\{1-(1+i)[C(\tau)-iS(\tau)]\},\quad \tau^{\frac{1}{2}}>0,$$

$$=e^{i\tau}\{1+(1+i)[C(\tau)-iS(\tau)]\},\quad \tau^{\frac{1}{2}}<0, \quad (69)$$

where $\tau^{\frac{1}{2}}[=(x-\bar{x})/\Omega]$ is positive for $r>\bar{r}$, and negative for $r<\bar{r}$. For large positive $p$ (in the shadow zone)

$$F(p)\rightarrow(3/2p)e^{-i\pi/6},\quad \int_0^\infty \exp[-i(2\tau^{\frac{1}{2}}t+t^2)]dt\rightarrow -i/2\tau^{\frac{1}{2}}, \quad (70)$$

and these leading terms just cancel each other in (67).

## 5. THE FIELD IN THE VICINITY OF THE HORIZON WHEN THE TRANSMITTER IS ON THE GROUND AND THE RECEIVER IS ELEVATED SEVERAL NATURAL UNITS OF HEIGHT ABOVE THE GROUND

When $h_1=0$, but $h>0$, we have $u_1=u_0$, and Eqs. (16), (15), and (20) yield

$$\Psi=[2e^{i\omega t}/(qk_0^2)^{\frac{1}{2}}]\int_0^\infty J_0(kr)kdku^{\frac{1}{2}}H_{1/3}^{(2)}(2u^{\frac{3}{2}}/3)/Q(k), \quad (71)$$

$$Q(k)\equiv[\beta u_0^{\frac{1}{2}}H_{1/3}^{(2)}(2u_0^{\frac{3}{2}}/3)-u_0H_{-2/3}^{(2)}(2u_0^{\frac{3}{2}}/3)]. \quad (72)$$

Proceeding now as in the previous section we obtain, in the case of strong absorption and a value of $z_2$ of several units,

$$\Psi\simeq\frac{2^{\frac{1}{2}}\exp[i\omega t+i\pi/4]}{(qk_0^2)^{\frac{1}{2}}(\pi r)^{\frac{1}{2}}}\int_{-i\infty}^\infty e^{-ikr}k^{\frac{1}{2}}dku^{\frac{1}{2}}H_{1/3}^{(2)}(2u^{\frac{3}{2}}/3)/Q(k), \quad (73)$$

---

[10] We are indebted to the Mathematical Tables Project for the use of their manuscript of Bessel functions of order ⅓.

[11] Jahnke-Emde, *Tables of Functions* (Dover Publications, New York, 1933), p. 35.

$$Q \simeq \beta u_0^{1/2} H_{1/3}^{(2)}(2u^{3/2}/3) = [(2/3^{1/2})\cdot(3v/2)^{1/3}]\cdot[iJ_{-1/3}(v)+e^{-i\pi/6}J_{1/3}(v)]\beta, \quad k<k_0,$$

$$= [(2/3^{1/2})\cdot(3w/2)^{1/3}]\cdot[iI_{-1/3}(w)-e^{-i\pi/6}I_{1/3}(w)]\beta, \quad k>k_0, \tag{74}$$

$$e^{-ikr}u^{1/2}H_{1/3}^{(2)}(2u^{3/2}/3) \rightarrow (3/\pi^{1/2})u^{-1/4}\exp[i5\pi/12-ip_5(k)], \quad p_5(k)=kr+(2u^{3/2}/3), \tag{75}$$

$$p_5(k)=p_5(k_0)-pv^{2/3}+s'v^{4/3}+\cdots=p_5(k_0)+pw^{2/3}+s'w^{4/3}+\cdots, \quad s'=(3/2)^{4/3}/4z_2^{1/2}, \tag{76}$$

$$\Psi \simeq \frac{\exp\{i[\omega t-(\pi/3)-k_0r-(2z_2^{3/2}/3)]\}(3\lambda/2\pi a_e)^{1/3}}{\pi(r\bar{r})^{1/2}\gamma(\epsilon_1-1)^{1/2}}G(p), \tag{77}$$

$$G(p)=\int_0^\infty \frac{\exp[-ipw^{2/3}]dw}{w^{1/3}[I_{-1/3}(w)+e^{i\pi/3}I_{1/3}(w)]}+\int_0^\infty \frac{\exp[ipv^{2/3}]dv}{v^{1/3}[J_{-1/3}(v)-e^{i\pi/3}J_{1/3}(v)]}$$

$$=\int_0^\infty \frac{\exp[-ipx^{2/3}]dx}{x^{1/3}[I_{-1/3}(x)+e^{i\pi/3}I_{1/3}(x)]}+e^{i\pi/3}\int_0^\infty \frac{\exp\{[(-3^{1/2}/2)+(i/2)]px^{2/3}\}dx}{x^{1/3}[I_{-1/3}(x)+e^{-i\pi/3}I_{1/3}(x)]}. \tag{78}$$

As before, the contribution from the $s'$ terms in the exponents can be estimated by using instead of $G(p)$, $[G(p)+is'(d^2G/dp^2)]$.

Values of $G(p)$ are given in Table IV, and $|G(p)|$ is plotted in Fig. 4. Figure 3 gives approximate values of $d^2G/dp^2$, which may be used for the purpose of estimating the correction due to the $s'$ term.

## 6. THE FIELD WHEN BOTH TRANSMITTER AND RECEIVER ARE ON THE GROUND (SURFACE WAVE)

The expression for the potential is obtained in this case by putting $u=u_0$ in (72):

$$\Psi=[2e^{i\omega t}/(qk_0^2)^{1/3}]\int_0^\infty J_0(kr)kdku_0^{1/2}H_{1/3}^{(2)}(2u^{3/2}/3)/Q(k). \tag{79}$$

Making use of the identity

$$\frac{u_0^{1/2}H_{1/3}^{(2)}(2u_0^{3/2}/3)}{Q}=\frac{(qk_0^2)^{1/3}}{[(k_0^2-k^2)^{1/2}+\gamma(k_1^2-k^2)^{1/2}]}\cdot\{-i+[u_0H_{1/3}^{(2)}(2u^{3/2}/3)-iu_0H_{-2/3}^{(2)}(2u_0^{3/2}/3)]/Q\}, \tag{80}$$

we can write

$$\Psi=\phi+\psi, \quad \phi=-2ie^{i\omega t}\int_0^\infty \frac{J_0(kr)kdk}{[(k_0^2-k^2)^{1/2}+\gamma(k_1^2-k^2)^{1/2}]}, \tag{81}$$

$$\psi=2e^{i\omega t}\int_0^\infty \frac{J_0(kr)kdk[u_0H_{1/3}^{(2)}(2u_0^{3/2}/3)-iu_0H_{-2/3}^{(2)}(2u_0^{3/2}/3)]}{[(k_0^2-k^2)^{1/2}+\gamma(k_1^2-k^2)^{1/2}]Q(k)}. \tag{82}$$

Here $\phi$ is Sommerfeld's potential of the surface wave for a *flat* earth, as can be seen by putting $h_1=h=0$ in (28). $\psi$ is an *additive* potential which represents the effect of the curvature of the earth; it vanishes in the limit of an infinite radius.

Using the same methods as in the preceding two sections we obtain, for the case of strong absorption,

$$\Psi \simeq \frac{\exp[i(3\pi/4)-irk_0]q}{(2\pi rk_0)^{1/2}(\epsilon_1-1)\gamma^2}H(p'), \quad p'=(3/2)^{2/3}(r/L), \tag{83}$$

$$H(p')=\int_0^\infty \exp[-ip'x^{2/3}]\cdot[E(x)-iF(x)]dx+\int_0^\infty \exp\{p'[(-3^{1/2}/2)+(i/2)]x^{2/3}\}\cdot[E(x)+iF(x)]dx, \tag{84}$$

$$E(x)-iF(x)=\frac{e^{-i\pi/6}[I_{-2/3}(x)-I_{1/3}(x)]+i[I_{-1/3}(x)-I_{2/3}(x)]}{[I_{-1/3}(x)+e^{i\pi/3}I_{1/3}(x)]}. \tag{85}$$

In the microwave region we may use for $\phi$ the leading term in the asymptotic expansion

$$\phi \rightarrow -\frac{2i\exp[-ik_0r]}{\gamma^2(\epsilon_1-1)k_0r^2}\left[1+\frac{i(3\epsilon_1+2)}{k_0r}+\cdots\right], \tag{86}$$

which follows by partial integration of van d. Pol's integral for $\phi$.[12] Combining (83) with (86) we obtain for the surface wave in a spherical earth

$$\Psi \simeq \frac{\exp[-i(\pi/2)+ik_0r]}{\gamma^2(\epsilon_1-1)k_0r^2}g(p'), \quad g(p')=1-\frac{2e^{i\pi/4}}{3\pi^{\frac{1}{2}}}(p')^{\frac{3}{2}}H(p'). \tag{87}$$

The function $H(p')$ has been computed from (84), and $|g(p')|$ is shown by the dashed curve in Fig. 5, where it is compared with a corresponding function $G(\eta)$ ($\eta=0.521p'$) computed by van d. Pol and Bremmer,[13] who used the normal mode theory.

## 7. DISCUSSION OF RESULTS

In this section we shall apply our theory to several cases which have been computed exactly by van d. Pol and Bremmer,[1] using the normal mode theory. The theory developed here is best suited for computing the field on the horizon and its immediate proximity. It becomes increasingly inaccurate as one recedes either far into the lit zone or far into the shadow zone. In addition, we also restricted ourselves to the case of strong absorption through the assumption that the quantity $\beta$ given in (21), (22), and (60) is large compared with unity. In the case of an elevated transmitter or receiver we have also assumed that the elevation is several units of natural height ($z_1$ or $z_2 \gg 1$), and that the quantities $s$, or $s'$ defined in (66) and (76) are small. All these assumptions become valid as the wave-length is decreased.

Figures 6 and 7 show a comparison of our results with exact values obtained by van d. Pol and Bremmer for an elevated transmitter and receiver. The case of a grounded transmitter is

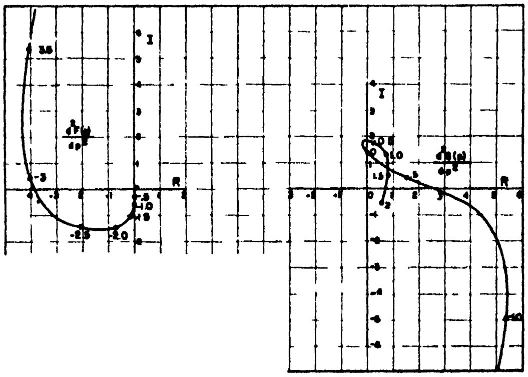


FIG. 3. Approximate values of $[d^2F(p)/dp^2]$ and $[d^2G(p)/dp^2]$. The numbers on the curves denote values of $p$.

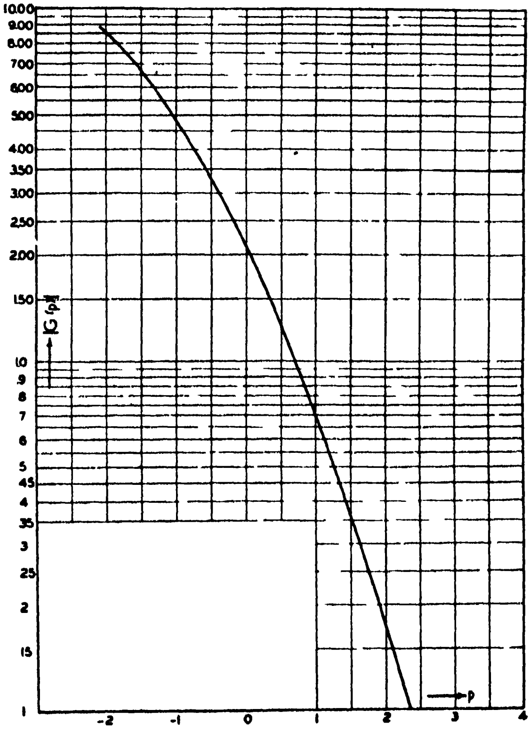


FIG. 4. The modulus of $G(p)$ defined in Eq. (78) and tabulated in Table IV.

---

[12] B. van d. Pol, Zeits. f. Hochfreqenztech. 37, 152 (1931).

[13] B. van d. Pol and H. Bremmer, Phil. Mag. 24, 173 (1937).

shown in Fig. 8. Some relevant quantities are tabulated in Table V, where the actual radius of the earth was used so as to conform to the calculations of van d. Pol and Bremmer. It is seen that in the cases treated the agreement near the horizon is sufficient for practical purposes for microwaves. Good agreement is also shown by Fig. 5 for the surface wave. Here our results do not go beyond those obtained by van d. Pol and Bremmer.

In a subsequent communication it will be shown that when the transmitter's height $z_1$ is small, equations (B) and (C) are to be multiplied by a factor $|(1+\beta z_1|$. When both $z_1$ and $z_2$ are small the potential $|\Psi|$ in (D) is to be multiplied by $|(1+\beta z_1)(1+\beta z_2)|$. Burrows and Gray[1] have computed a function $F(L)=(3^{\frac{1}{2}}/2\pi)\times|G[(8/9)^{\frac{1}{3}}p]|$, using the normal mode theory. These authors also give $F(L)$ for the case of a perfectly conducting earth.

## APPENDIX

## The Amplitude of the Reflected Wave

If we integrate the second term in (32) around the stationary point $k^*$ of $p_3(k)$ defined in (44) we obtain, in a manner similar to the derivation of (51), for the amplitude $\Psi_1'$ of the reflected wave

$$\Psi_1'=\frac{\exp[i\omega t-ip_3(k^*)]R(k^*)(qx)^{\frac{1}{2}}}{(rP)^{\frac{1}{2}}(1+qh_1-x^2)^{\frac{1}{4}}\cdot(1+qh_2-x^2)^{\frac{1}{4}}}. \tag{a}$$

$$\begin{aligned}P\equiv 2(1+qh_1-x^2)^{\frac{1}{2}}-[2x^2/(1+qh_1-x^2)^{\frac{1}{2}}]\\+2(1+qh_2-x^2)^{\frac{1}{2}}-[2x^2/(1+qh_2-x^2)^{\frac{1}{2}}]\\-4(1-x^2)^{\frac{1}{2}}+[4x^2/(1-x^2)^{\frac{1}{2}}];\quad x=k^*/k_0.\end{aligned} \tag{b}$$

Here $R(k)$, the reflection coefficient for a plane earth, is given in (26). In deriving it we used the asymptotic expression (25), which is valid only if $u_0$ is large. This is

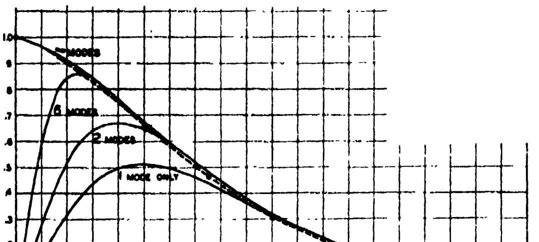


FIG. 5. Comparison of our results for the surface wave with those of van d. Pol and Bremmer (Phil. Mag. 24, 173, 1937). ------ $|g(p')|$, Eq. (87); ——— the corresponding function from normal mode theory using various modes.

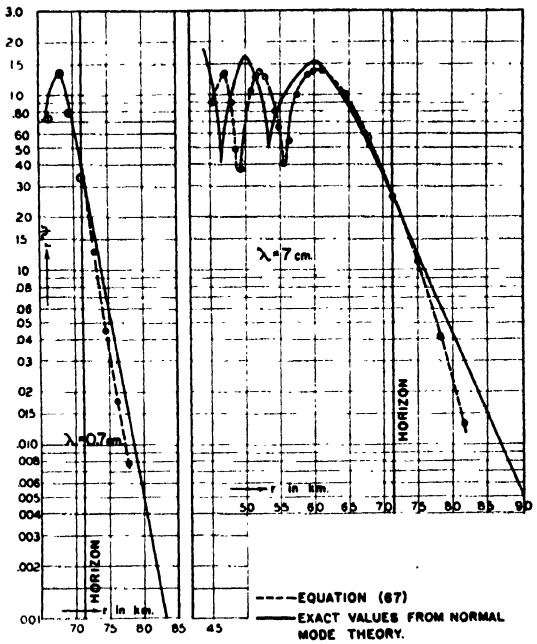


FIG. 6. Comparison of our results with those of van d. Pol and Bremmer.[2] Transmitter and receiver at a height of 100 m. $\epsilon=4$, $\sigma=10^{-13}$ e.m.u.

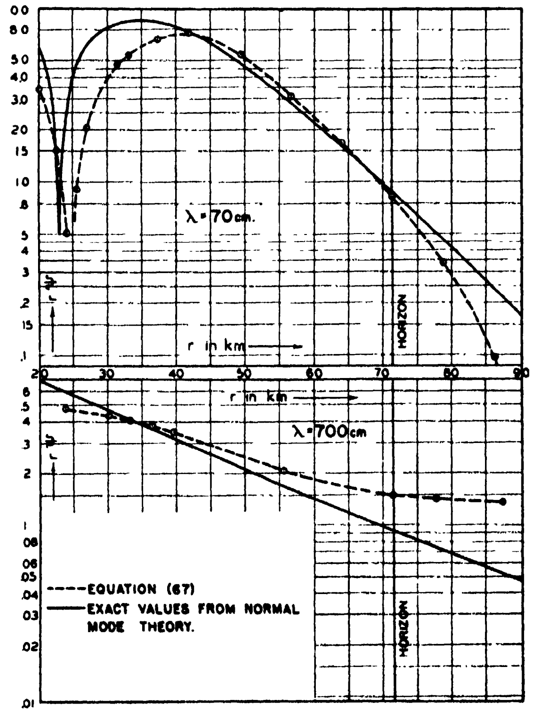


FIG. 7. Comparison of our results with those of van d. Pol and Bremmer.[2] Transmitter and receiver at a height of 100 m. $\epsilon=4$, $\sigma=10^{-13}$ e.m.u.

not true near grazing incidence, where $u_0$ becomes small. For microwaves, however, the spherical reflection coefficient is still well represented by $R(k)$ even at grazing incidence, except in the limiting case of a perfectly conducting ground.[11]

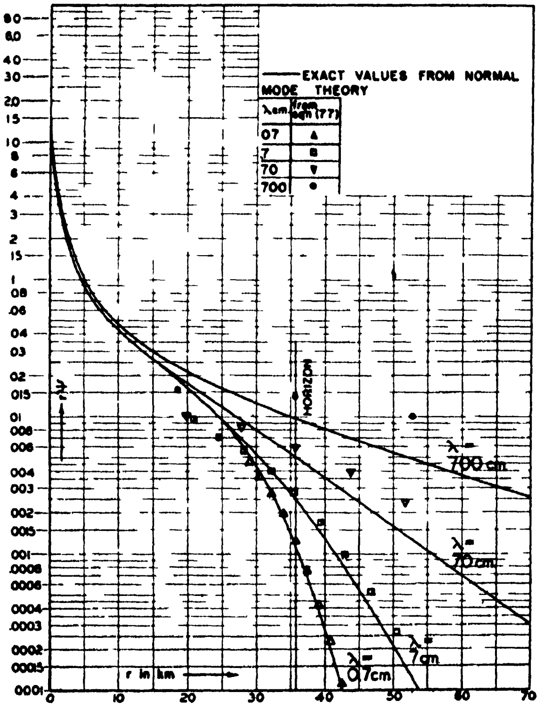


FIG. 8. Comparison of our results with those of van d Pol and Bremmer.[1] Transmitter on the ground. Receiver at a height of 100 m. $\epsilon=4$, $\sigma=10^{-18}$ e.m u.

TABLE V. Some quantities pertaining to the cases treated by van d. Pol and Bremmer,[1] $\epsilon=4$, $\sigma=10^{-18}$ e.m.u.

| $\lambda$ | $s_1$ or $s_2$ | $\varsigma$ | $\varsigma'$ | $\lvert\beta_v\rvert$ |
|---|---|---|---|---|
| 0.7 cm | 63 | .11 | .054 | 614 |
| 7 cm | 14 | .23 | .12 | 285 |
| 70 cm | 2.9 | .50 | .25 | 132 |
| 700 cm | .63 | 1.08 | .54 | 56 |

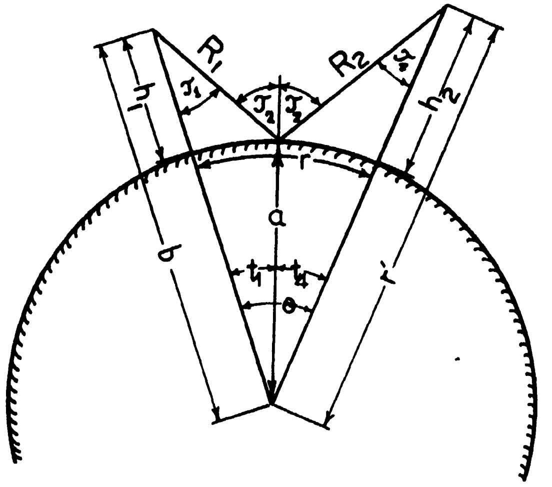


FIG. 9. The ray reflected from the surface of the earth

We shall now compare (a) with the expression for the reflected wave derived by van d. Pol and Bremmer,[15] and which includes the so-called divergence factor:

$$\Psi_1'=\frac{\exp[\imath\omega t-\imath p_3(k^*)]R(k^*)a(\sin\tau_2\cos\tau_2)^{\frac{1}{2}}}{[br'\sin\theta(R_1r'\cos\tau_4+R_2b\cos\tau_1)]^{\frac{1}{2}}},\tag{c}$$

where the quantities are defined in Fig. 9. We have, to within terms of the order $q^2h^2$,

$$\begin{gathered}1+qh_1=b^2/a^2;\quad 1+qh_2=r^2/a^2,\quad q=2/a,\\ (1+qh_1-x^2)^{\frac{1}{2}}=(b/a)\cos\tau_1,\quad (1+qh_2-x^2)^{\frac{1}{2}}=(r'/a)\cos\tau_4,\\ x=\sin\tau_2,\ R_1=b\cos\tau_1-a\cos\tau_2,\ R_2=r'\cos\tau_4-a\cos\tau_2,\end{gathered}\tag{d}$$

$$P=2(R_1\cos\tau_4\cos t_1+R_2\cos\tau_1\cos t_4)/(a\cos\tau_1\cos\tau_2\cos\tau_4).\tag{e}$$

Substituting in (a) we get

$$\Psi_1=\frac{\exp[\imath\omega t-\imath p_3(k^*)]a(\sin\tau_2\cos\tau_2)^{\frac{1}{2}}}{[br'r(R_1\cos\tau_4\cos t_1+R_2\cos\tau_1\cos t_4)]^{\frac{1}{2}}}.\tag{f}$$

Since $\sin\theta\simeq r/a$, and $\cos t_1$ and $\cos t_4$ are close to unity, expressions (c) and (f) agree within the region of validity of the earth-flattening approximation.

---

[14] Reference 2, p. 268.

[15] Reference 13, p. 844, and reference 2, p. 266.

# Letters to the Editor

## Linearized Supersonic Flow through Ducts

H. F. Ludloff and F. Reiche
*College of Engineering, New York University, New York, New York*

April 22, 1947

In various aerodynamical problems, it is desirable to be able to compute the supersonic field of flow through ducted bodies by an analytic, approximate method, which yields quick insight into essential flow features as related to the shape of the nozzle wall. The linearized theory of supersonic flow suggests solutions that can be produced by spreading "supersonic sources" of various intensities over the surface of a hollow skeleton cylinder. If the source density is assumed constant over this surface, the wall contour of a converging diverging nozzle results. With a source intensity which increases linearly along the generatrices of the cylinder, one obtains ducts of gradually decreasing cross section. In the first case, one obtains a duct with a dull rim, and at the entrance a compression shock which has the shape of a double Mach cone, subtended over the rim as the base. This shock is weak near the rim, but becomes stronger toward the vertex and remains particularly strong over the whole surface of the second Mach cone. The linear source distribution produces a knife-sharp rim, as a consequence of which the shocks can be shown to disappear. A certain combination of these two source distributions hinders the formation of shocks, and produces a converging diverging nozzle, whose inside diameter increases at the exit section to the same magnitude as it had at the entrance section. In this way no wave drag results.

Derivation of these results is based on computing the velocity potential from which the velocity and stream function can be derived:

$$\varphi=-2d\cdot\int_0^{\omega^*}d\omega\int_0^{x-\beta R}d\xi\cdot\frac{f(\xi)}{[(x-\xi)^2-\beta^2R^2]^{\frac{1}{2}}}.$$

Here $f(\xi)$ is intensity of source layer at point with coordinates $\xi$ (along generatrix) and $\omega$ (azimuth) on surface of cylinder; $x-\xi$ and $R$, axial and radial components of pseudo-distance between source point and point under consideration. $R^2=d^2+r^2-2d\cdot r\cdot\cos\omega$; $d$ radius of cylinder, $r$ radial distance between axis and point under consideration.

$$\omega^*=\cos^{-1}\frac{d^2+r^2-x^2/\beta^2}{2d\cdot r};\quad \beta^2=M^2-1.$$

(I). For constant source density ($f(\xi)=A$), we get:

$$\varphi=-2d\cdot A\cdot\int_0^{\omega^*}d\omega\cdot\log\frac{\beta R/x}{1-[1-(\beta R/x)^2]^{\frac{1}{2}}}.\tag{1}$$

Differentiating and introducing $R^2=\rho$ instead $\omega$ as integration variable:

$$v_r=\frac{\partial\varphi}{\partial r}=\frac{xd\cdot A}{\beta r}\int_b^c\frac{d\rho}{\rho}\cdot\frac{(\rho-(a\cdot b)^{\frac{1}{2}})}{[(c-\rho)(a-\rho)(\rho-b)]^{\frac{1}{2}}},\tag{2}$$

where

$$a=(d+r)^2,\quad b=(d-r)^2,\quad c=x^2/\beta^2.$$

The expression for $v_r$ can be discussed.

(a) at $x/\beta=d-r$, i.e. at front cone, by expansion:

$$v_r=-\pi\cdot A\cdot\left(\frac{d}{r}\right)^{\frac{1}{2}}\left\{1-\frac{3d+r}{4d}\left[\frac{x/\beta-(d-r)}{2r}\right]\cdots\right\},\tag{2a}$$

which shows discontinuity, $-\pi\cdot A\cdot(d/r)^{\frac{1}{2}}$, of $v_r$ at front cone.

(b) at $x/\beta=d+r$, i.e. at second Mach cone, by forming:

$$\lim_{c\to a}v_r\sim\lim_{c\to a}\frac{xd\cdot A}{\beta r}\int_b^c\frac{d\rho}{\rho}$$
$$\times\frac{\rho-(a\cdot b)^{\frac{1}{2}}}{[(c-\rho)(a-\rho)(\rho-b)]^{\frac{1}{2}}}\to\log[(x/\beta)^2-(d+r)^2],\tag{2b}$$

which shows coalescence of singularities and logarithmic infinity of $v_r$, at second Mach cone.

(c) in the far rear region, by expanding after $\beta R/x$:

$$v_r=\frac{\pi\beta^2r\cdot d\cdot A}{x^2}\left\{1+\frac{3}{4}\frac{\beta^2}{x^2}(r^2+2d^2)\cdots\right\},\tag{2c}$$

which shows that disturbance velocity vanishes asymptotically.

The stream function can be obtained by means of Stokes three-dimensional representation:

$$\psi=-\rho_0\int_{(d-r)\cdot\beta}^{x}r\cdot\frac{\partial\varphi}{\partial r}\cdot dx'.\tag{3}$$

Upon inserting (2) into (3), we can discuss $\psi$, upon integrating (2a), (2b), and (2c) with respect to $x$. In this way one can see that $\psi$

(a) has no discontinuity at front cone, since it is proportional to $[x/\beta-(d-r)]$,

(b) is regular and finite at second Mach cone, since it is an integral over a logarithmic infinity,

(c) vanishes asymptotically as $1/x$.

The streamtubes are computed from the relation $U\cdot\rho_0\cdot(r^2/2)+\psi(x,r)=U\cdot\rho_0\cdot(d^2/2)$; they show a constriction between first and second Mach cone near the entrance, and at infinity an inside diameter of the same magnitude as at the entrance.

(II). For linearly increasing source distribution ($f(\xi)=\bar{A}\cdot\xi$), we get:

$$\varphi=2\bar{A}\cdot d\int_0^{\omega^*}d\omega\left\{-x\cdot\cosh^{-1}\frac{x}{\beta\cdot R}+(x^2-\beta^2\cdot R^2)^{\frac{1}{2}}\right\}\tag{1}$$

and

$$\frac{\partial\varphi}{\partial r}=v_r=\frac{\bar{A}\cdot\beta\cdot d}{r}\int_b^c\frac{d\rho}{\rho}\frac{(c-\rho)^{\frac{1}{2}}(\rho-(a\cdot b)^{\frac{1}{2}})}{[(a-\rho)(\rho-b)]^{\frac{1}{2}}}.\tag{2}$$

Therefore, (a) at $x/\beta=d-r$, i.e. at front cone:

$$v_r\sim\frac{x/\beta-(d-r)}{(d\cdot r)^{\frac{1}{2}}}\xrightarrow{\pi}0,\tag{2a}$$

which means that there is no discontinuity of $v_r$ at entrance, i.e. knife-edged rim.

(b) at $x/\beta=d+r$, i.e. at second Mach cone: $v_r$ remains finite; in (II), (2) $(c-\rho)^{\frac{1}{2}}$ appears now in numerator; therefore for $c\to a$ $(c-\rho)^{\frac{1}{2}}$ and $(a-\rho)^{\frac{1}{2}}$ cancel, rather than forming a cumulative singularity.

(c) at far rear, expansion after $\beta R/x$ yields:

$$v_r=-\frac{\bar{A}\pi\cdot\beta^2r\cdot d}{x}\left\{1+\frac{1}{4}\frac{\beta^2}{x^2}(r^2+2d^2)+\cdots\right\}.\tag{2c}$$

As before, the stream function is obtained by integration with regard to $x$, so that at far rear:

$$\psi=\rho_0\bar{A}\pi\beta^2r^2\cdot d\cdot[\log x+\cdots].\tag{3c}$$

Asymptotically, $\psi$ becomes logarithmic infinite; the cross section of the duct decreases as $(r/d)^2 \sim 1/(1+c\cdot\log x)$. For practical purposes, the flow remains parallel within a wide range of $x$, but a considerable wave drag results.

The advantages of both source distributions (absence of shocks; wide exit section) can be combined, by assuming linearly increasing source intensity $\bar{A}\cdot\xi$ from the rim to a distance $\xi=B$, from where a constant source intensity $\bar{A}\cdot B$ continues. Such a source combination can be resolved into a linear source distribution, starting at the rim, and a linear sink distribution, starting a distance $B$ behind. It is easy to see that the source layer produces a conical compression shock, while the sink layer starts a subsequent expansion wave of equal strength, so that no strong shock will develop anywhere and no wave drag result.

---

# Rigorous Solution of Linearized Supersonic Flow Through Ducts

C. E. Mack, Jr., *Grumman Aircraft Engineering Corporation*
AND
H. F. Ludloff and F. Reiche, *College of Engineering, New York University, New York*

April 22, 1947

In the case of constant source strength along a skeleton cylinder, a rigorous solution of the linearized problem of flow through ducted bodies may be obtained in closed form. The method of approach is such that complete elliptic integrals of the first, second, and third kind arise; however, complete integrals of the third kind may be expressed in terms of elliptic integrals of the first and second kinds so that existing tables suffice for numerical computation.

As in the case of the series expansions, the solution assumes different forms depending upon the location of the test point $P(x, r)$. If $P$ is so located that its fore cone cuts some but not all of the generatrices of the skeleton cylinder, then $P$ is said to lie in the front district. If the fore cone of $P$ cuts all the generatrices of the skeleton cylinder, then $P$ lies in the rear district. Each region will be considered separately.

### FRONT DISTRICT $\beta(d-r)<x<\beta(d+r)$

If the value of $v_r$ as given by Eq. (I, 2) above is substituted into Eq. (I, 3) for the disturbed stream function, the latter is given by

$$\psi=-\frac{\rho_0 d\cdot A\beta}{2}\int_{c'=b}^{c'=c}dc'\int_{\rho=b}^{\rho=c'}\frac{d\rho}{\rho}\frac{(\rho-(ab)^{1/2})}{[(c'-\rho)(a-\rho)(\rho-b)]^{1/2}} \quad (1)$$

where the symbols have the same meaning as previously. This integral may best be handled by reversing the order of integration, *viz:*

$$\psi=-\frac{\rho_0 d\cdot A\beta}{2}\int_{\rho=b}^{\rho=c}\frac{d\rho(\rho-(ab)^{1/2})}{\rho((a-\rho)(\rho-b))^{1/2}}\int_{c'=\rho}^{c'=c}\frac{dc'}{(c'-\rho)^{1/2}}. \quad (2)$$

Carrying out the integration with respect to $dc'$,

$$\psi=-\rho_0 d\cdot A\beta\left[\int_{\rho=b}^{\rho=c}\frac{(c-\rho)^{1/2}d\rho}{((a-\rho)(\rho-b))^{1/2}}\right.$$
$$\left.-(ab)^{1/2}\int_{\rho=b}^{\rho=c}\frac{(c-\rho)^{1/2}d\rho}{\rho[(a-\rho)(\rho-b)]^{1/2}}\right]. \quad (3)$$

If now the substitution $\rho=b+(c-b)\sin^2\theta$ is made, then after a short reduction the value of $\psi$ is easily seen to be

$$\psi=-2\frac{\rho_0 d\cdot A\beta}{(a-b)^{1/2}}[(a-b)E_1(k)-(a-c-(ab)^{1/2})F_1(k)$$
$$-c(a/b)^{1/2}\Pi_1(n,k)], \quad (4)$$

where $F_1$, $E_1$, and $\Pi_1$ are complete elliptic integrals of the first, second, and third kinds, the parameters $k$ and $n$ being given by

$$k^2=(c-b)/(a-b),\quad 0\leq k<1;\quad n=(c-b)/b,\quad 0\leq n<\infty.$$

It has been shown by Cayley that the complete elliptic integral of the third kind may be written, in our notation, as

$$\left[\frac{(n+1)(n+k^2)}{n}\right]^{1/2}\Pi_1(n,k)=\frac{\pi}{2}+\left[\frac{n+k^2}{n(n+1)}\right]^{1/2}F_1(k)$$
$$+F_1(k)F(k',\theta)-F_1(k)E(k',\theta)-E_1(k)F(k',\theta), \quad (5)$$

where $F(k',\theta)$, $E(k',\theta)$ are incomplete elliptic integrals the parameters of which are

$$k'^2=1-k^2,\quad \cot^2\theta=n.$$

When the value of $\Pi_1$ from Eq. (5) is substituted into Eq. (4) and the dimensionless ratio $Y=(r/d)(0\leq Y<1)$ introduced, it follows at once that

$$\psi=-\rho_0 d^2\cdot A\beta\{4Y^{1/2}[E_1(k)-k'^2F_1(k)]$$
$$-[(1-Y)^2+4k^2Y]^{1/2}[\pi+2\{F_1(k)-E_1(k)\}F(k',\theta)$$
$$-2F_1(k)E(k',\theta)]\}. \quad (6)$$

The stream lines may be obtained from this by evaluating the right-hand side for a suitable range of parameters, combining the results with the stream function for the undisturbed stream and interpolating to find the curves along which the total $\psi$ is constant.

In a similar way to that used above and by the same substitution $\rho=b+(c-b)\sin^2\theta$ it is easy to find the velocity components $v_x=\partial\varphi/\partial x$ and $v_r=\partial\varphi/\partial r$. The results are

$$v_x=-\frac{4Ad}{\beta(a-b)^{1/2}}F_1(k), \quad (7)$$

$$v_r=\frac{2d\cdot A(c)^{1/2}}{r(a-b)^{1/2}}\left[F_1(k)-\left(\frac{a}{b}\right)^{1/2}\Pi_1(n,k)\right]. \quad (8)$$

The elliptic integral of the third kind may be removed as before.

### REAR DISTRICT $\beta(d+r)<x<\infty$

The determination of the disturbed stream function for the rear region is very similar to that of the front region, so only the final result need be stated. The value of $\psi$ is given by

$$\psi=-\rho_0 d^2\cdot A\beta\left\{\frac{4}{k}Y^{1/2}\left[E_1(k)-\frac{1-Y}{1+Y}k'^2F_1(k)\right]\right.$$
$$-\left[(1-Y)^2+\frac{4Y}{k^2}\right]^{1/2}[\pi+2\{F_1(k)-E_1(k)\}F(k',\theta)$$
$$\left.-2F_1(k)E(k',\theta)]-4Y^{1/2}+2(1+Y)\tan^{-1}\frac{2Y^{1/2}}{1-Y}\right\}, \quad (9)$$

where now the parameters $k$, $k'$, $\theta$ are different than for the front region. We have here

$$k^2=(a-b)/(c-b),\quad 0\leq k<1;\quad \cot^2\theta=(a-b)/b$$

and of course $k'^2 = 1 - k^2$. The velocity components are essentially the same as before and need not be reproduced.

Although the solution above is for the case of constant source strength, it may be shown that for polynomial source distributions, the stream lines may be obtained by repeated integration of the stream lines for the constant source. For example, the stream lines for a linear source distribution may be obtained by a single graphical or numerical integration from the constant source stream lines.

---

## The Mechanism of Cutting Fluid Action

MILTON C. SHAW
*Mechanical Engineering Department, Massachusetts Institute of Technology, Cambridge, Massachusetts*

April 23, 1947

IN a recent letter to the editor,[1] Mr. G. P. Brewington questioned the observation of Ernst and Merchant[2] concerning the lack of correlation between the efficiency of a metal cutting fluid and its physical properties. In support of his contention that some relationship does exist between the cutting ratio and the physical properties of the fluid, Mr. Brewington cites the fact that materials having certain end groups present in their chemical structure (such as $-SH$ or $-Cl$) give a high cutting ratio. It is not clear how this observation substantiates Mr. Brewington's point of view. While it is true that the nature of the end group present on a molecule has an influence on its tendency to become adsorbed on a metal surface, the possibility of the cutting fluid reacting chemically with the metal cut should not be overlooked.

Considerable experimental data have been collected in recent years which indicate that certain effective metal cutting fluids do act by reacting chemically with the metal during the cutting operation. These observations support a rather simple theory of cutting fluid action in which the fluid reacts chemically in the neighborhood of the tool point to form a low shear strength solid lubricant which, in turn, prevents metal-to-metal contact between chip and tool. This mechanism of cutting fluid operation has been described[3] as being a chemico-physical action inasmuch as the fluid first reacts to form chemical products which, in turn, prevent metal-to-metal contact by a physical separation of the surfaces.

The afore-mentioned correlation between chemical structure and cutting performance is, from the writer's point of view, evidence of the correctness of the chemico-physical mechanism of cutting fluid action. Direct chemical analysis has established the presence of solid chemical products of low shear strength on the surface of metal cut in the presence of effective cutting fluids. The fact that certain good cutting fluids are also effective when applied to the tool in their vapor form, the tool and work piece being maintained at a temperature in excess of the boiling point of the fluid, is further evidence of the importance of the chemical rather than physical nature of the fluid. Other observations such as the dependence of the effectiveness of a cutting fluid upon the nature of the metal cut and the fact that benzene is a poorer cutting fluid than air can be explained in terms of the chemical aspects of the problem.

Probably the strongest argument in favor of the chemico-physical theory of cutting fluid action is offered by the discovery that metal cutting processes may be used in the manufacture of chemical products. In this case the product is a chemical rather than a finished machine part. Certain conditions which are found to exist at the point of a cutting tool appear to have an important bearing upon the initiation of a chemical reaction. Three such conditions are:

1. high local pressure—a pressure which will approach the hardness of the metal cut;
2. high local temperature—a temperature which will locally approach the melting point of the metal cut;
3. a nascent highly reactive and highly stressed clean metal surface.

These three conditions may be used to advantage in many organo-metallic reactions particularly those reactions which are difficult to start. In such a case the work piece becomes the metal reactant while the cutting fluid takes the form of the liquid reactant. This method of carrying out a chemical reaction has been termed "mechanical activation."[4] While mechanical activation represents an example of cutting fluid action in which the chemical products formed at the interface between the chip and the tool are present on a macroscopic scale, and, indeed, in some instances represent the entire quantity of metal cut, it should be evident that this technique is in reality a limiting case of cutting fluid action.

The writer concurs in Mr. Brewington's belief that the normal primary monohydric alcohols derive their limited effectiveness as cutting fluids for aluminum from chemical action, and has previously published[5] some substantiating data.

The physical properties of a cutting fluid which control the rate and quantity of fluid which can find its way to the cutting point might well influence the over-all performance of a metal cutting fluid. There thus appear to be two distinct phases of the general problem. First, the physical properties of a useful fluid must be such as to insure sufficient mobility that a satisfactory quantity of fluid reach the cutting point. Second, after the fluid has reached the tool point it is necessary that its chemical nature be such as to enable it to react chemically with the metal cut to form a low shear strength solid lubricant. The required mobility of a cutting fluid will depend largely upon the cutting speed, becoming of negligible significance at very low speed. The afore-mentioned tests of Ernst and Merchant were carried out at a very low cutting speed (5.5 inches per minute) and hence no correlation between cutting efficiency and the physical properties of the cutting fluid should be expected. However, if high speed cutting data were to be examined with regard to the physical properties influencing the mobility of the fluid (viscosity and surface tension), some correlation might be found.

[1] G. P. Brewington, "On the Action of Cutting Oils," J. App. Phys. **18**, 260 (1947).

[2] H. Ernst, and M. E. Merchant, "Chip Formation, Friction and High Quality Machined Surfaces," paper in "Symposium on Surface Treatment of Metals," Am. Soc. Metals, Cleveland (1940), pp. 299–335.

[3] M. C. Shaw, "The Chemico-Physical Role of the Cutting Fluid," Met. Prog. **42**, 85 (1942).

[4] M. C. Shaw, "Mechanical Activation," USP 2,416,717 (1947).

[5] M. C. Shaw, "Action of *n*-Primary Alcohols as Metal Cutting Fluids—Alternating Properties with Chain Length," J. Am. Chem. Soc. **66**, 2057 (1944).

# New Books

## A Chronological History of Electrical Development

NATIONAL ELECTRICAL MANUFACTURERS ASSOCIATION, 155 East 44 Street, New York, 1946. 106 pp. Price $2.

This is an historical record of electricity and the devices and equipment created by the electrical manufacturing industry to utilize it; of a long line of men of science, their initiative, theories, research, experiments, and discoveries, and their contributions to the present age of electricity, electronics, and atomic energy. The book is designed as a reference work for editors, writers, commentators, libraries, and schools, as well as for industry.

While the book touches on such modern electrical applications as television, fluorescent and ultraviolet light, the giant turbo-generator, and the electron micro-analyzer for atomic research, its accent is on the early-day men of science and industry and the steps they took to make electrical living possible. Other contributing factors are recorded, among them electrical expositions and congresses, the birth of electrical publications, engineering and trade associations, societies, foundations, and institutes.

## Methods of Mathematical Physics

BY HAROLD JEFFREYS AND BERTHA SWIRLES JEFFREYS. Pp. 679+VII. Cambridge, at the University Press; The Macmillan Company, New York, 1947. Price $15.00.

As the authors state in their preface, this book is intended to provide an account of those parts of pure mathematics that are most frequently needed in physics. The manner in which they propose to carry out their treatment is revealed by a statement in the preface that careful analysis is more important in science than in pure mathematics, not less, and that the easiest way to make a statement reasonably plausible is to give a rigorous proof. It is this reviewer's feeling that that intention has been carried out in ample measure and in most adequate style.

Chapter 1 treats the real variable and includes such topics as the Bolzano-Weierstrass (without so naming it) and Heine-Borel theorems, Riemann and Stieltjes integrals, uniform continuity and uniform convergence, and infinite integrals. Chapters 2 and 3 deal with vectors and tensors, respectively, the treatment throughout bearing in mind the student of physics and emphasizing such topics as the strain and stress tensors, Euler's dynamical equations, parallax, and the motion of a top. Chapter 4 is on matrices and contains a treatment of quadratic and Hermitian forms, the Pauli spin matrices, the Eddington and Dirac 4×4 matrices, crystal structure, and electromagnetic theory. In Chapter 5, we find multiple integrals, including line and surface integrals, Green's and Stokes' theorems, flux and circulation. A careful analysis is made of the notions of area and of length. Potential theory is the content of Chapter 6. Poisson's and Laplace's equations, minimal theorems, the Rayleigh-Ritz method of numerical solution are treated in detail. Operational methods and physical applications of the operational method occupy Chapters 7 and 8, respectively. The applications deal with charging of a condenser, the seismograph, resonance, dissipative and gyroscopic systems, and radioactive disintegration. Chapter 9 (numerical methods) includes such topics as Lagrange's interpolation formula, divided differences, the central difference formula, formulas for the value of a definite integral, numerical solution of algebraic and differential equations. A note on graphical methods recommends that they be honored in the breach rather than in observance. They are best avoided entirely, the authors advise us. Chapter 10 treats the calculus of variations and leads up to the Hamilton equations, principle of least action, and Routh's modified Lagrangian function. Chapters 11, 12, and 13 deal with functions of a complex variable and their application to contour integration and conformal representation. The Joukowsky transformations are treated in detail. Fourier series are the subject matter of Chapter 14. Approximation by polynomials, as well as by trigonometric series, is discussed, also the Laplace transform. Chapter 15 is given over to the factorial and related functions. Our authors frown on the gamma-function as a matter of notation and prefer to label it the factorial function. Wallis's formula for the value of $\pi$ and Stirling's formula for the value of $\log Z!$ are derived. Chapter 16 is devoted to second-order linear differential equations. These are studied as regards ordinary and singular points and are solved in series and by complex integrals. Infinite determinants are touched on in this connection. Asymptotic expansions occupy Chapter 17, with application to such matters as wave velocity, dispersion of water waves, and refraction of a pulse. Potential waves and heat conduction, waves in one dimension and spherical symmetry, conduction of heat in one and three dimensions, and allied physical ideas are treated in Chapters 18, 19, and 20. The many important systems of coordinates are treated in detail. Bessel functions in Chapter 21 and applications of Bessel functions in Chapter 22 are a logical continuation of the subject matter in the three preceding chapters. In Chapter 23 are found the confluent hypergeometric function, its representation in series and by complex integrals, Whittaker's transformation, Hermite and Laguerre polynomials, and Schrödinger's equation for the hydrogen-like atom. Chapter 24 is given over to Legendre and associated functions and their application to potential theory. Chapter 25, on elliptic functions, closes the text.

Thirteen pages of notes follow, numbered according to the sections of the text to which they refer and designed to clarify or amplify such sections. For example, the Heine-Borel theorem, alluded to in Section 1.072 of the text, is proved in the corresponding section of the notes.

The appendix on notation, which completes the book, embodies recommendations on standardizing certain bits of notation—particularly those with regards to $\nu$ and $\varphi$—which, as is well known, have not been used with uniformity of meaning in the literature.

The legends posted below the chapter headings are often delightful and illuminating. To quote a few: "one by one, or all at once" (multiple integrals), "but all that moveth doth mutation love" (potential theory), "even Cambridge mathematicians deserve justice" (operational methods), "I have no satisfaction in formulas unless I feel their numerical magnitude" (numerical methods).

Each chapter is followed by a list of problems, in all—if the reviewer's count is correct—230, and culled, in a large measure, from the British collections such as the Mathematical Tripos.

A high degree of thoroughness, such as will satisfy the pure mathematician, with the entire work set in a frame of physical applications, which enliven and give meaning to the mathematics treated, make this book a welcome addition to the literature of mathematical physics.

MAX MORRIS
*Case School of Applied Science*

## Photography by Infrared. Its Principles and Applications

BY WALTER CLARK. Second Edition. Pp. 472+xvii, Figs. 93, 21½×14 cm. John Wiley and Sons, Inc., New York, 1946.

The London air raids destroyed the printing plates for the first edition of this book, which was published in England in 1939. The author has taken full advantage of this opportunity to revise thoroughly and modernize the presentation. Included are not only many wartime developments both here and abroad, but also more fundamental advances such as the improvement of infra-red sensitizing dyes.

The result is a comprehensive and authoritative account of the subject, and one that will appeal particularly to the amateur photographer, to the scientific research worker, and to the technician interested in possible application of infra-red. The author states that some acquaintance with photographic practice is assumed, although a fair amount of space is devoted to ordinary photographic procedures, with only occasional reference to the modifications required for work in the infra-red.

This material, including a chapter on darkroom practice, one on the characteristics of films, and an appendix on developers, etc., was presumably included to enhance the value of the book as a laboratory guide. However, it also furnishes a good background for discussion of the problems peculiar to the infra-red, and the general reader with little experience in photography will appreciate the complete and logical way in which the subject as a whole is presented.

The introductory chapter is mostly historical, and the inclusion of direct quotations and illustrations from the early publications, both here and elsewhere in the book, is an attractive feature. Incidentally, it is unfortunate that the very first figure, depicting Newton's experiment on the dispersion of white light, is incorrectly drawn. The reviewer was pleasantly surprised to find due credit given to the physicists and astronomers who did the pioneer work in infra-red photography, because all too often this work has been overlooked. The second chapter contains a useful summary of the general practice of infra-red photography, with references to later chapters where the various matters are treated in more detail. This ingenious arrangement, while very suitable for the purpose in hand, leads to a considerable amount of repetition.

In keeping with the objective of discussing principles as well as applications, further chapters deal with the sensitometric characteristics of photographic materials, sensitization for the infra-red, indirect methods on infra-red photography, and sources of infra-red. These subjects are well presented, and embody a wealth of valuable information. The only criticism that might be made is the failure to distinguish clearly at the outset between the effects of the intensity of the light and of the exposure time. Thus "exposure" is used as abscissas for the characteristic curves without stating that it is the product of intensity and time, and the situation is not improved by giving the wrong units in the first place in which they appear. A very good discussion of the failure of the reciprocity law is given later, but the author has followed the almost universal tendency of writers in this field first to give the impression that the two variables are equivalent, and afterwards to correct this impression.

The remainder of the book is devoted to the various applications of infra-red photography, chief among which are differentiation of materials (textiles, printer's inks, paintings, documents), medicine, botany and paleontology, photo-micrography, landscape photography, forest and other surveys, special effects photography, photography in the dark, spectrographic and astronomical photography, and camouflage detection. Penetration of haze and fog is covered both from the theoretical and observational standpoints, and here the author rightly emphasizes that, contrary to the popular impression, infra-red has little advantage where the large particles of natural fogs are concerned. A final chapter deals with the optical properties of materials in the infra-red. It includes data on the transmission and reflection of common materials, on infra-red filters, and an especially valuable discussion of the infra-red focus of lenses.

Rarely does one find a book as well documented as this one. The bibliographies following each chapter frequently contain over one hundred titles, in many languages. Although a few of the latest developments, such as the concentrated arc and the interference filter, are not included, the coverage of the subject is remarkably complete. This, together with the abundance of valuable data and the many fine illustrations, should justify the prediction that the book will be the standard reference work in infra-red photography for some time to come.

J. A. JENKINS
*University of California*

## Introduction to Electron Optics. The Production, Propagation and Focusing of Electron Beams

BY V. E. COSSLETT. Pp. 272, Figs. 155, 16½×25 cm. The Clarendon Press, Oxford, England. 1946.

*Introduction to Electron Optics* is primarily a textbook based on a series of lectures delivered by the author, at Oxford, as part of a course in electron optics. The book maintains a reasonably good balance between mathematical manipulation, physical ideas, and experimental results. The treatments of the various topics are not so long or detailed as to result in a ponderous book but still include sufficient material to serve as a good introduction to the general subject.

The subject matter is presented in the following order, with the first five topics taking up about two-thirds of the total space:

1. Electrostatic fields
2. Electrostatic focusing
3. Magnetic focusing
4. Aberrations
5. Production of electron beams
6. Cathode-ray tubes
7. Electron diffraction and electron microscopes
8. Miscellaneous application.

The mathematical treatments of lenses can be readily followed by students with training in mathematics through differential equations and vector analysis. The presentation is clear and no difficulty with the meanings or symbols should be encountered. (A list of mathematical symbols is appended.) The discussions are well illustrated by line drawings and graphs, as well as numerical data.

The book is intended for students of electron optics, although those doing research in this field will likely find it useful. Each chapter contains a list of references for further reading along the lines of the material of the particular chapter. The use of a general bibliography of a few carefully chosen references for each phase of the subject is to be commended.

The section dealing with electron diffraction is very brief as is that concerned with electron microscopy. In these two instances the reference might have been somewhat more extensive, but it is very likely that those who are interested in applied electron optics will have access to other sources. The emphasis is on fundamentals rather than design and use.

In the opinion of the reviewer, this book should serve exceedingly well as a basis for courses in electron optics in American universities.

R. D. HEIDENREICH
*Bell Telephone Laboratories, Murray Hill, New Jersey*

## Relaxation Methods in Theoretical Physics

BY R. V. SOUTHWELL. Pp. 248. Figs. 117, 16×24½ cm. Oxford, at the Clarendon Press, Oxford, England, 1946.

The technique of solving numerically the Laplace equation in one dimension, by reduction to a difference equation and subsequent application of "relaxation" networks, has become a well-known procedure in advanced problems in engineering. Professor R. V. Southwell and his colleagues have contributed notably to this advance in applied mathematics. Their work has been summarized in exceedingly useful form in Professor Southwell's book, *Relaxation Methods in Engineering Science*, published in 1940. It is highly gratifying, therefore, to have at hand a companion volume which treats in a similar way the two-dimensional problems of mathematical physics.

The book under review, *Relaxation Methods in Theoretical Physics*, is devoted to the solution of the partial differential equation $\frac{\partial}{\partial x}\left[\chi\frac{\partial\psi}{\partial x}\right]+\frac{\partial}{\partial y}\left[\chi\frac{\partial\psi}{\partial y}\right]+Z=0$, where $\chi$ and $Z$ may be functions of $x$ and $y$, and $\psi(x, y)$ is the function sought, subject to prescribed boundary conditions. The arrangement of the book is excellent, and presents enough detail so that the interested reader may understand the background theory and carry through the computational techniques to a useful end.

Two chapters are devoted to the development of the approximating difference equations for the case $\chi=$ constant in a form for numerical solution and to the relaxation techniques for square, triangular, and hexagonal networks. Practical hints for the beginner are included. As pointed out by the author, one can gain a clear insight into the method only by doing the examples in detail.

Irregular networks applied to various boundary conditions are considered in Chapter III. Three types of boundary conditions are treated, *viz.*, (1) $\psi$ has a given value on the boundary, (2) $\partial\psi/\partial\nu$, the normal derivative of $\psi$, has specified values on the boundary, and (3), a combination of these two conditions. In many cases boundaries of complex shape may be reduced to circular or square boundaries by conformal transformation so that computation by simple nets is more readily effected.

Problems of more complexity, in which $\chi$ is a function of $x$ and $y$, are discussed in detail in Chapter V. While problems of this type are more difficult to solve by orthodox methods than those of the plane harmonic type ($\chi=$constant), they may be solved numerically with only slight modification in the relaxation technique.

The power of the relaxation method becomes fully apparent in its application to problems in which different equations govern the wanted function or in which some part of the boundary conditions must be determined by computation. Such problems are next to insoluble by orthodox methods. The last chapter of the book treats in detail physical problems of the above type.

An appendix contains formulae for numerical integration and differentiation, a bibliography of papers relating to relaxation methods, a list of the problems solved as examples in the text, and large charts illustrating the solutions. These include torsion problems, conformal transformations, oil pressure distributions in bearings, temperature distribution in a piston, percolation of liquids through walls or strata of various types, etc.

In conclusion one may mention that the basic aim of the numerical approach to physical problems, so admirably treated in this book, is to provide a flexible solution commensurate with the accuracy of the experimental data involved. The wide utility of the relaxation method should be made known to all serious students of physics and engineering who are concerned with more than a "handbook" solution to physical problems. In providing these two basic books, Professor Southwell has performed a major service for the advancement of mathematics in engineering.

S. W. MCCUSKEY
*Case Institute of Technology*

## The Mathematical Tables Project

Sponsored by the National Bureau of Standards. Columbia University Press, New York.

In a recently published book, *An Index of Mathematical Tables*, three British authors have erected a monument to those scientists, who from 1596 on, have published the tables reviewed in the index, an estimated one thousand tables. While it is not possible to contemplate this vast array of published tables without experiencing a feeling of gratitude to the many computers whose work is represented, at the same time it is to be observed that the tables reviewed are published in a bewildering assortment of ways. Some are actually unavailable, and many are practically so. Some are to be found buried in periodical literature, and others constitute a few pages in scattered books. Some of the tables appear as isolated volumes, and still others appear in series of volumes.

The fact that the publications of the New York Project appear as a sequence of tables of similar size and binding, and are sold at nominal prices, makes the series of incalculable value to the modern scientist. As the volumes appear, the set becomes more and more the first source to be consulted when a mathematical table is wanted. The writer of this review looks forward to the day when any table of general usefulness will be found among these volumes.

The care that is exercised in the preparation of these tables, in order to prevent the printing of wrong values, can neither be briefly described nor readily appreciated by any but the professional computer.

The writer of this review regrets that the bound volumes of this table have not been imprinted with distinguishing numbers. In the back of each volume is placed a complete list of all of the volumes prepared prior to the date on which that volume was published. In this list each title bears a reference number. These numbers make better references to a particular volume than do the rather long and similar titles. It seems unfortunate that this number does not appear on the outside cover nor even on the title page. A complete list of titles follows in which those titles which have been recently received are starred.

(1) Table of the First Ten Powers of the Integers from 1 to 1000. 80 pp. (1939). Out of print.
(2) Tables of the Exponential Function $e^x$. XV+535 pp. (1939).
(3) Tables of Sines and Cosines for Radian Arguments. XIX+275 pp. (1940).
(4) Tables of Circular and Hyperbolic Sines and Cosines for Radian Arguments. XVII+405 pp. (1939).
(5) Tables of Sine, Cosine, and Exponential Integrals, Vol. I. XXVII+444 pp. (1940).
(6) Tables of Sine, Cosine, and Exponential Integrals, Vol. II. XXXVII+225 pp. (1940).
(7) Table of Sine and Cosine Integrals for Arguments from 10 to 100. XXXII+185 pp. (1942).
(8) Tables of Probability Functions.
Vol. I. XXVIII+306 pp. (1941) Vol. II. XXI+348 pp. (1941).
(9) Table of Natural Logarithms.
Vol. I. XVIII+505 pp. (1941) Vol. II. XVIII+505 pp. (1941).
Vol. III. XVIII+505 pp. (1941) Vol. IV. XVIII+510 pp. (1941).
(10) Miscellaneous Physical Tables. VII+61 pp. (1941).
(11) Tables of the Moment of Inertia and Section Modulus of Ordinary Angles, Channels, and Bulb Angles, with Certain Plate Combinations. XIII+201 pp. (1941).
(12) Table of Arc tan$x$. XXV+173 pp. (1942).
* (13) Table of Circular and Hyperbolic Tangents and Cotangents for Radian Arguments. XXXVIII+412 pp. (1943).
* (14) Table of the Bessel Functions $J_0(z)$ and $J_1(z)$ for Complex Arguments. XLIV+406 pp. (1943).
* (15) Table of Reciprocals of the Integers from 100,000 through 200,009. VIII+204 pp. (1943).
* (16) Tables of Lagrangian Interpolation Coefficients. XXXVI+394 pp. (1944).
* (17) Table of Arc sin$x$. XIX+124 pp. (1945).
* (18) Tables of Associated Legendre Functions. XLVI+306 pp. (1945).
* (19) Tables of Fractional Powers. XXX+490 pp. (1946).
* (20) Tables of Spherical Bessell Functions. Vol. I. XXVIII+378 pp. (1947).

ORLEY E. BROWN
*Case Institute of Technology*

## Wave Propagation in Periodic Structures. Electric Filters and Crystal Lattices

BY LEON BRILLOUIN. Pp. 247, 5×8½ in. McGraw-Hill Book Company, Inc., New York, 1946. Price $4.00.

The rapidly increasing rate at which scientific knowledge is accumulating at the present time would appear to forecast an age of greater and greater specialization in the sciences. Such a state of affairs must undoubtedly be viewed with apprehension by many workers in these domains who have derived a great part of the pleasure in their work from the broadness of understanding with which scientific pursuits have supplied them in the past, and who believe that the interplay of ideas from various specialized fields of a subject such as physics is a healthy and stimulating influence on the development of the subject as a whole. That these fears are as yet ungrounded is demonstrated with force and clarity by Professor Brillouin in this small volume of less than 250 pages, wherein a surprisingly small body of mathematical discipline and associated physical ideas are shown to be quite adequate for the description of a wide range of seemingly diverse physical phenomena, running the gamut from mile long electrical transmission lines to the passage of electrons through crystal lattices.

As its title indicates, the book deals with the general aspects of wave propagation through periodic structures, and examples and applications are drawn from many fields in which such phenomena occur: electric and mechanical filters, vibrations of crystal lattices, x-ray diffraction in crystals, and the quantum theory of metals, to quote some leading examples. The characteristic appearance of transmission and attenuation bands for waves propagating in such periodic structures is demonstrated to have its important physical consequences in each of the above phenomena in such a way that one realizes that a comprehension of the phenomena in any one case almost automatically leads to the understanding of the principles involved in the others. The mathematical apparatus required to discuss adequately almost all of the problems which arise is of such simplicity as to require only the barest acquaintance with the calculus as background for understanding the development of the ideas presented.

This volume, unlike many books, is not intended as a reference work but as a key to understanding. No attempt is made to treat in detail the various phenomena selected as examples for the general ideas presented. These may be found in many other books. But there are few, if any, other books which demonstrate so directly the broader comprehension which can be obtained of these same physical phenomena when one steps back and examines them as a group linked by a common mathematical and physical background. It is in this respect that the true value of the book becomes immediately apparent.

Professor Brillouin is well known through his other books for the clarity of his exposition, and it is indeed fortunate that he has chosen to make available to others his thoughtful understanding of a subject to which he has been an outstanding contributor.

LESLIE L. FOLDY
*University of California*

# New Booklets

North American Philips Company, Inc., 100 East 42 Street, New York 17, New York, has published a 6-page folder (R1063) entitled *Industrial Control with X-Ray Diffraction*. The text describes the principles of x-ray diffraction and how it is performing difficult jobs in modern industrial applications. Typical x-ray diffraction film is shown and explained. Geiger-counter x-ray spectrometer is discussed in connection with qualitative and quantitative analysis work.

North American Philips also announces a new 4-page folder describing the new Norelco air trimmer, which consists of a stator having three concentric cylinders that slide in the spaces between the four concentric rings of the rotor. Free on request.

New Enterprises, Inc., 84 State Street, Boston 9, Massachusetts, has issued a brochure setting forth its objectives and methods of operation. It was incorporated in February, 1946, "for the purpose of making venture capital more readily available for the development of untried business enterprises of a scientific nature. . . . To qualify for consideration, a new venture must be one of pronounced technical characteristics--a new process, a new or improved technique which shows promise of significant future possibilities."

Leeds and Northrup Company, 4902 Stenton Avenue, Philadelphia 44, Pennsylvania, issues quarterly a $10\frac{1}{2}\times15\frac{1}{2}$-inch bulletin containing numerous photographs and articles describing the use of Leeds and Northrup controls in industry. 8 pages.

Fischer and Porter Company, Department 6F-C, Hatboro, Pennsylvania, has published a new bulletin describing its "Flowrator" instruments for measuring flow rate of liquids and gases. 24 pages. Available on request.

*Bakelite Review* for April 1947 contains 28 pages of attractive pictures and articles about consumer goods molded of plastic material. Published quarterly by Bakelite Corporation, 30 East 42d Street, New York 17, New York.

The Tube Department of Radio Corporation of America, Harrison, New Jersey, has resumed issuance of *Application Notes*, which were popular before the war. The first postwar Note is called *Input Admittance of Receiving Tubes*. 20 pages.

Another new publication of the Tube Department is Form No. CRPS-102 entitled *RCA Phototubes, Cathode-Ray Tubes, and Special Tubes*. 16 pages, 10 cents per copy.

Public Affairs Committee, Inc., 22 East 38th Street, New York 16, New York, has published a 32-page pamphlet entitled *The Struggle for Atomic Control*. It was written by William T. R. Fox, Associate Professor of Political Science at Yale University. 20 cents per copy.

Corrosion Publishing Company, 1131 Wolfendale Street, Pittsburgh 12, Pennsylvania, is announcing a booklet of 94 pages entitled *Stress Corrosion Cracking of Mild Steel* by James T. Waber and Hugh J. McDonald. A monograph, it is in large part a reprint of a series of articles which appeared in *Corrosion and Material Protection*. It presents a general theory of stress corrosion and the importance of nitrogen was deduced from it. $2 per copy.

Farrand Optical Company, Inc., Bronx Boulevard and East 238th Street, New York 66, New York, has published *Bulletin 801* entitled *Ultra-Violet Monochromators*. Six instruments are pictured and described. 12 pages.

Burrell Technical Supply Company, 1942 Fifth Avenue, Pittsburgh 19, Pennsylvania, offers a 4-page leaflet describing its precision Mixblend laboratory mixer, *Baker's Analyzed* C. P. Potassium Iodide, Taylor slide comparators.

General Aniline and Film Corporation, 270 Park Avenue, New York 17, New York, has issued a new *Carbonyl Iron Powder* booklet. 34 pages. Available on request.

International Resistance Company, 401 North Broad Street, Philadelphia 8, Pennsylvania, announces its *Bulletin C-2*, which gives complete specifications, characteristics, and dimensions for IRC fixed and adjustable power wire wound resistors, ferrule, and bracket types. 6 pages. Available on request.

*The General Radio Experimenter* for March, 1947, published by General Radio Company, Cambridge, Massachusetts, features an article on search receivers for radar counter-measures. 8 pages. Available on request to engineers, scientists, and technicians writing on company or professional letterhead.

# Here and There

## Personnel

Charles C. Bidwell, head of the department of physics at Lehigh University, retired July 1 with the rank of Professor Emeritus.

Five new members have been appointed to the technical staff of the Midwest Research Institute, as follows: in inorganic chemistry, Solon D. Fisher; in organic chemistry, Kenneth R. Hoffman; in engineering mechanics, Eugene F. Hamilton, Richard H. Scheil, and Tage Mortensen.

J. Robert Oppenheimer of the University of California has been appointed to succeed Frank Aydelotte in October, 1947, as Director of the Institute for Advanced Study. Dr. Oppenheimer is a graduate of Harvard and received advanced training in Cambridge University and Göttingen, where he took his Ph.D. He has for the last decade been professor of theoretical physics both in the University of California and in the California Institute of Technology. He was director of the laboratory at Los Alamos which perfected the atomic bomb and is a member of various Government committees for the control of atomic energy.

The Institute for Advanced Study has three Schools: Mathematics, which includes mathematical physics; Economics, which includes history and political science; and Humanistic Studies, which consists at the moment largely of Greek archaeology and the history of art. The Institute provides facilities for postdoctorate research. The Doctor's degree is required for admission, and no degrees are given by the Institute itself.

The appointment of Professor Oppenheimer does not indicate any change in the policy of the Institute as regards the subjects which will be pursued. As it happens, Professor Oppenheimer is a scientist who had as an undergraduate a broad humanistic training.

Dr. Aydelotte emphasized very strongly his approval of this appointment. It would have been difficult to find in the United States a man more ideally qualified for the directorship.

The Institute has a strict rule of retirement at sixty-five. Dr. Aydelotte was kept on until sixty-seven by the trustees at the urgent request of the faculty. He will continue to occupy offices at the Institute for his work in connection with the Rhodes Scholarships, the Guggenheim Fellowships, and the American Friends Service Committee.

## U. S. Atomic Energy Commission Announces Distribution of "Heavy Water"

Heavy water and deuterium gas are now being made available by the United States Atomic Energy Committee for research purposes within the United States. The abundance of the heavy hydrogen isotope in the material is approximately 99.9 percent. Quantities will be limited to normal research requirements.

The Stuart Oxygen Company of San Francisco, California, will act as contracting agent for distribution. This company began small-scale production and distribution soon after the discovery of deuterium. In the postwar period, because of the increased need for deuterium in research the Stuart Company's production has been far below the demand. The quantities being made available are from stock produced by other operators during the war for the Manhattan District.

The material will be distributed with charges based on the cost of handling and distribution. The cost of production of the material itself will not be included. Distribution in this manner is being effected under authority of the Atomic Energy Act of 1946 which provides for the fostering and assistance of research by the Commission.

Allocation will be handled in a manner similar to that for radio-isotopes. For complete information write to U. S. Atomic Energy Commission, Research Division, Field Operations, Oak Ridge, Tennessee, attention Isotopes Branch.

## Special Expedition Studied Eclipse

A B-29 was used by the National Geographic Society-Army Air Forces cooperative expedition on May 20 in connection with the total eclipse of the sun in Brazil. The plane carried a heavy load of scientific instruments for making studies of cosmic rays in the vicinity of the equator.

Other tests performed by the cooperative expedition dealt with Einstein's theory of relativity, under the direction of Professor G. Van Biesbroeck, astronomer of the Yerkes Observatory of the University of Chicago. The principal objective was an effort to determine whether light rays from certain stars are bent as the rays pass near the sun on their way to the earth, as predicted by Einstein in his theory of relativity.

As a result of the influence of the gravitational field of the sun, light is bent, and star images registered on the photographic plates appear to be shifted away from the sun slightly beyond their actual positions. If this shift amounts to 1.75 seconds of arc, it will be additional evidence that Einstein's theory is correct.

As a check on how much the star images are moved on his photographic plates, Professor Van Biesbroeck will photograph the same stars at night six months later from the same spot. At that time the sun will not be present to bend their rays. Measuring the difference, if any, between the positions of the star images on the two photographs taken six months apart will show how much the light rays were bent during the eclipse.

## New Observatory at Colorado

The University of Colorado is collaborating with Harvard University on the creation of a new solar observatory devoted to the study of the region of space which lies between the sun and earth. The new station will be an expansion of the present High Altitude Observatory, where Dr. Walter O. Roberts carries on a study of the sun's outer regions with a coronagraph.

According to present plans, the heart of the new set-up will be a large tube, about six feet in diameter and twenty feet long, set in a dome. In the tube will be placed at least five observing instruments: a large coronagraph, a small coronagraph to provide automatic "tracking" on the sun, another large telescope devoted to observation of short-wave radiation, and two smaller ones to keep continuous moving picture records of routine sun conditions. In addition, other instruments will aid in the study of cosmic rays and paths of meteorites as a means of providing information about the outer reaches of the atmosphere.

### Polymer Research Clinics at Brooklyn

Two summer laboratory clinics remain in the series sponsored by the Institute of Polymer Research of the Polytechnic Institute of Brooklyn. From July 21 to 26 a clinic will be held on "Advanced X-Ray Diffraction". Attendance will be limited to about ten, and the fee is $100. Professor I. Fankuchen is in charge. The September 8–12 clinic is on the subject of "Polymerization Techniques". Attendance will be limited to twelve; fee is $150. Professor Herman Mark is in charge.

### Microwave Fellowships at Stanford

The Microwave Laboratory, Division of the Physics Department of Stanford University, has announced a number of fellowships, research assistantships, research associateships, and postdoctorate research fellowships for the academic year 1947–1948. Students who are working in the Microwave Laboratory may receive their degrees in either the Physics or Electrical Engineering Departments. Stipends for the four classes of awards range from $1000 to $4000. For further information and application address Professor William W. Hansen, Director, Microwave Laboratory, Stanford University, California.

### Matthew Luckiesh Awarded I.E.S. Medal

Matthew Luckiesh, director of the Lighting Research Laboratory, General Electric Company, has been awarded the I.E.S. Gold Medal, given annually by the Illuminating Engineering Society "for meritorious achievement which has conspicuously furthered the profession, art, and knowledge of illuminating engineering." The medal will be presented to Dr. Luckiesh at the Illuminating Engineering Society National Technical Conference to be held September 15–19 in New Orleans.

### Symposium on Applied Mathematics

A Symposium on Applied Mathematics will be held at Brown University, Providence, Rhode Island, August 2–4, 1947, under the sponsorship of the American Mathematical Society and the cosponsorship of the American Institute of Physics. The subject of the Symposium will be "Nonlinear Problems in the Mechanics of Continua." All persons interested in receiving further information concerning the Symposium should write to Professor W. Prager, Brown University, Providence 12, Rhode Island. Copies of the program and reservation cards will be sent to them three weeks before the Symposium.

## Calendar of Meetings

| | |
|---|---|
| **July** | |
| 11–12 | American Physical Society, Stanford University, California |
| 16–19 | American Society of Civil Engineers, Duluth, Minnesota |
| **August** | |
| 2–4 | American Mathematical Society (Symposium on Applied Mathematics), Brown University, Providence, Rhode Island |
| 21–23 | Society of Automotive Engineers, Los Angeles, California (Transportation and Maintenance Meeting) |
| 26–29 | American Institute of Electrical Engineers, San Diego, California |
| **September** | |
| 1–2 | Mathematical Association of America, New Haven, Connecticut |
| 1–4 | American Society of Mechanical Engineers, Salt Lake City, Utah |
| 8–12 | Instrument Society of America, Chicago, Illinois |
| 15–19 | American Chemical Society, New York, New York |
| 15–19 | Illuminating Engineering Society, New Orleans, Louisiana |
| 17–18 | Society of Automotive Engineers, Milwaukee, Wisconsin |
| 24–27 | Conference on Electrical Insulation, Cambridge, Massachusetts |
| **October** | |
| 2–4 | Society of Automotive Engineers, Los Angeles, California |
| 15–18 | American Society of Civil Engineers, New Orleans, Louisiana |
| 20–24 | American Society for Metals, Chicago, Illinois |
| 21–25 | American Chemical Society, California Section, San Francisco, California |
| 23–25 | Optical Society of America, Cincinnati, Ohio |
| **November** | |
| 3–7 | American Institute of Electrical Engineers, Chicago, Illinois |
| 6–7 | Society of Automotive Engineers, Tulsa, Oklahoma (Fuels and Lubricants Meeting) |
| 28–29 | American Physical Society, Houston, Texas |
| **December** | |
| 1–3 | Society of Automotive Engineers, Kansas City, Missouri (Air Transport Engineering Meeting) |
| 1–5 | American Society of Mechanical Engineers, New York, New York |
| 26–31 | American Association for the Advancement of Science, Chicago, Illinois |
| 29–31 | American Physical Society, Chicago, Illinois |

Journal
of
Applied Physics

Volume 18, Number 8 August, 1947

# Infinite-Rejection Filters

ALBERT M. STONE AND JAMES L. LAWSON
*Radiation Laboratory,* Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Received May 20, 1946)

The characteristics of certain bridged-$T$ filter structures have been investigated theoretically and an experimental confirmation of this theory at u.h.f. has been obtained. In theory it is possible to obtain infinite attenuation at a designated frequency, while retaining essentially the same band width as that for the uncompensated filter. A model has been constructed which has a band width of $\frac{1}{2}$ Mc/sec. at 3000 Mc/sec. with an attenuation at center frequency of over 70 db. This is to be compared with the attenuation from a similar uncompensated filter of 20 db. Certain other known filter expressions have also been reduced to forms readily utilizable at these frequencies. Their utility for u.h.f. applications is discussed. The distortion produced in otherwise rectangular pulses of short duration has been investigated both theoretically and experimentally. Curves are presented which show the resultant wave form as a function of the tuning of the filter.

## I. INTRODUCTION

It has been known for some time that a parallel-resonant filter of the bridged-$T$ structure may be designed for complete (i.e., infinite) rejection of the frequency to which it is tuned. This infinite rejection is obtained regardless of the finite $Q$ of the coils and condensers of the filter and is limited solely by one's skill in adjusting a single resistive element.[1] At the same time as long as the ratio of unloaded[2] $Q$ to loaded $Q$ (*viz.*, line loading plus external resistance loading) is large compared with unity, the band width, measured between the points where the power transmitted is 3 db below the off-resonance value, is virtually indistinguishable from that of the normal (i.e., finite rejection) filter consisting of the tuned circuit alone. The principal object of this article is to present these results in readily usable form for u.h.f. work and to point out some salient features hitherto unstressed.

The applications of infinite-rejection filters are as varied as ingenuity affords. The particular problem which prompted the present analysis was that of eliminating an interfering *c-w* (or low-frequency modulated) carrier without, at the same time, destroying the usefulness of a pulsed 10-cm radar system. The analysis, therefore, is divided into two parts: (1) an investigation of the criteria for infinite rejection at microwave frequencies, and (2) an investigation of the distortion produced in rectangular pulses by such filters.

---

* This paper is based on work done for the Office of Scientific Research and Development under Contract OEMsr-262 with the Massachusetts Institute of Technology.

[1] W. N. Tuttle, Proc. I.R.E. **28**, 23, 1940; H. W. Bode, U. S. Patent No. 2,002,216 (1933).

[2] The $Q$ of a bandpass circuit will be defined as the ratio of the center frequency to the band width. For a simple circuit such as the single-tuned type, $Q = L\omega_0/r$, where $r$ is the series resistance, $L$ the inductance, and $\omega_0$ the resonant radian frequency.

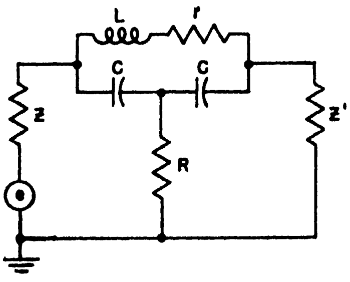


FIG. 1. Schematic diagram of an infinite-rejection filter with tap between condensers.

The essential modification of the ordinary filter consists of a tap from the center of the coil or condenser arm to ground (see Fig. 1) through a resistance $R$. (There may be reactance in this arm as well; it has been disregarded, however, as it merely complicates matters and leads to nothing fundamentally new.) It will be shown in Section II that a proper choice of $R$ (see Fig. 1) can be made so that *no* power is passed from the generator to the load at the resonant frequency. In fact the resistance required is $R=(rQ_0^2)/4$, where $Q_0$ is the unloaded $Q$ of the circuit, defined as $L\omega_0/r$, $\omega_0$ is the resonant radian frequency, and the other symbols have conventional meanings. It is immaterial whether the coil or the condenser is center-tapped. However, if the circuit is considered to be an $L$, $R''$, $C$ all in parallel, a center tap on the resistance $R''$ brought to ground through $R$ will not achieve the desired result. Physically this must be so, for at resonance the parallel $L$, $R''$, $C$ circuit is merely $R''$. A circuit of this type obviously cannot attenuate completely (see Fig. 2).

In the third section an alternative method of accomplishing this same result will be presented which, however, can be shown to be precisely equivalent to the methods mentioned above. The application of these results to *i-f* and video frequencies will appear obvious. Several possible methods for obtaining infinite-rejection u.h.f. filters will be discussed in this section and some experimentally obtained results will be presented.

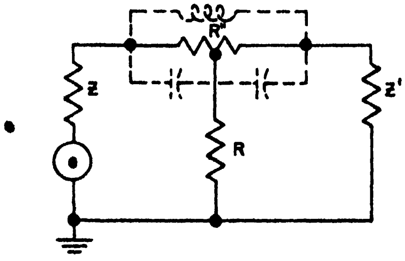


FIG. 2. Modification of Fig. 1 not leading to infinite rejection.

Finally, in Section IV the results of an analysis of the pulse distortion produced by such a filter will be presented.

## II. CIRCUIT ANALYSIS

The circuit of Fig. 1 is in the form of a bridged-$T$ section with the capacitance tapped and hence can be developed by the usual methods of circuit analysis[3] into the symmetrical-lattice form, Fig. 3. This lattice is precisely equivalent to the bridged-$T$, having the same poles and zeros and the same frequency behavior. For simplicity the generator and its internal impedance (or the line impedance if we consider the generator matched to the line) and the load have been split off. The dotted lines signify repetition of the corresponding elements. It is not difficult to see that the symmetrical lattice is nothing more nor less than the usual four-arm bridge, as shown in Fig. 4, where $Z_a$ is the series element and $Z_b$ is the shunt element of the lattice. Now from this it is easy to picture physically why the circuit of Fig. 1 gives infinite rejection. To balance the bridge one tunes the resonant circuit as is customary, and then, by adjusting $R$, a condition of zero load current is obtained. In such a state the power that would normally pass on to the load is absorbed in the added resistance. This condition of balance is, obviously, the one that makes $Z_b=Z_a$. It is unnecessary to carry this calculation out in detail. The result may be

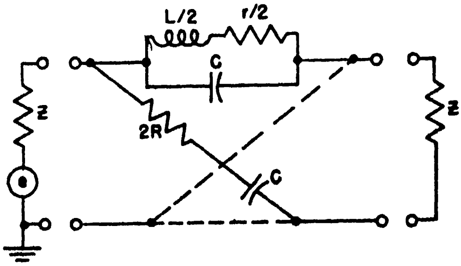


FIG. 3. Lattice development of infinite-rejection filter.

---

[3] E. A. Guillemin, *Communication Networks* (John Wiley and Sons, New York, 1935), Vol. 2, Chap. IV.

stated; the bridge is balanced when

$$\omega_0^2 = 2/LC, \tag{1}$$

and, simultaneously,

$$R = (rQ_0^2)/4. \tag{2}$$

Equation (1) states that the tuning of the circuit is independent of the added shunt element (but this is only true if it is non-reactive). Equation (2) gives the added condition that increases the maximum attenuation in the resonance curve indefinitely. It is clear that, for the ordinary values of $Q_0$ met in high frequency and u.h.f. work, $R$ is easily realizable physically. If the two capacitances in Fig. 1 are unequal, the same conditions (1) and (2) hold, but $C/2$ is now the series sum of the two.

The case of the tapped coil leads to results which are expressed precisely by Eq. (1) and Eq. (2). (Compare Fig. 5 and Fig. 1.) The symmetrical-lattice development of the tapped-coil circuit is shown in Fig. 6. If the two half-coils are identical, the value of $L$ to be used in Eq. (1) is

$$L = 2L_1 + 2kL_1, \tag{3}$$

where $L_1$ is the self-inductance of the half-coil, and $k$ is the coefficient of coupling. (This choice of identical coils is not an essential restriction but leads to facility in computation.)

It is desirable for ease of discussion, once the criterion for infinite rejection has been established, to compute the actual output power passed by the filter as a function of frequency; or, rather, to compute the dimensionless fraction, the output power divided by the output power in the absence of the filter. The latter output power is simply $e^2/4Z$, if it is assumed that load and line are matched and that $e$ represents the generator voltage. It is sufficient to consider only the circuit of Fig. 1. Let $i_L$ be the output current. We form the dimensionless quantity $F$ by

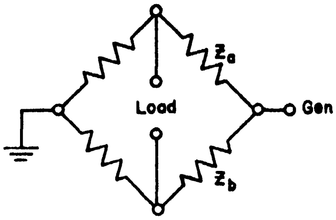


Fig. 4. Circuit equivalent to Fig. 3.

$$F = \frac{4|i_L|^2 Z^2}{e^2}. \tag{4}$$

We are not interested in any phase factor since the output of the filter usually is applied to the grid of a tube. The following definition is introduced:

$$\omega/\omega_0 \equiv m = \frac{\omega(LC)^{\frac{1}{2}}}{2}.$$

Then, with the previous definition of $Q_0$ and the balance condition, Eq. (2), one arrives at

$$F = \frac{4Z^2R^2[m^2Q_0^2(m^2-1)^2 + (1-m^2)^2]}{(2R+Z)^2\left[mZQ_0(m^2-1) - \dfrac{4mR}{Q_0}\right]^2 + \left[2RZ + \dfrac{4R^2}{Q_0^2} - m^2(Z^2+6RZ+4R^2)\right]^2}. \tag{5}$$

In simplifying this expression it is convenient to introduce the further definition

$$R \equiv \beta Z/4. \tag{6}$$

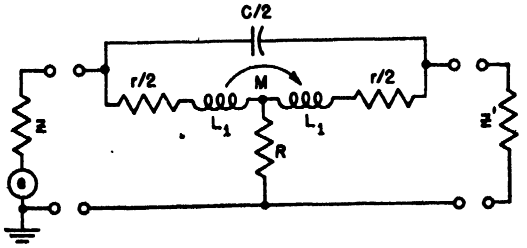


Fig. 5. Infinite-rejection filter with tapped coil.

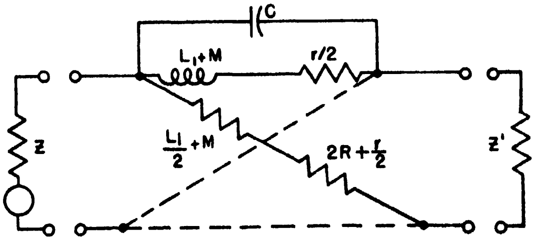


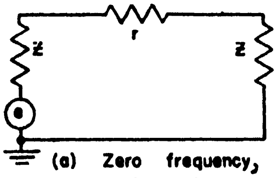
Fig. 6. Lattice development of Fig. 5.

Thus $\beta$ is the *standing-wave ratio at resonance looking toward the unbalanced filter terminating* *the line.*[4] (In this connection use is made of the fact that, for a reasonably high-$Q$ filter, $L\omega_0/r = R'/L\omega_0$, where $R'$ is the equivalent shunt resistance.) If terms in $1/Q^4$ and higher powers in the denominator of Eq. (5) are neglected, that equation can be written

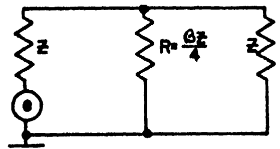
$$F \cong \left[\frac{\beta}{\beta+2}\right]^2 \left[\frac{m^2(1-m^2)^2}{m^2(1-m^2)^2+\left(\frac{\beta+2}{2Q_0}\right)^2}\right]. \quad (7)$$

For most practical purposes $m=1$ very closely. Hence Eq. (7) becomes

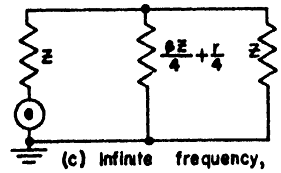
$$F \cong \left[\frac{\beta}{\beta+2}\right]^2 \frac{(1-m^2)^2}{(1-m^2)^2+\left[\frac{\beta+2}{2Q_0}\right]^2}. \quad (8)$$



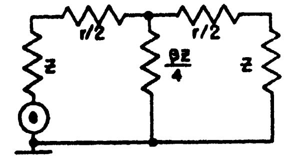
(a) Zero frequency, tapped condenser



(b) Infinite frequency, tapped condenser



(c) Infinite frequency, tapped coil



(d) Zero frequency, tapped coil

FIG. 7. Limiting cases of Figs. 1 and 5.

In the region far from resonance, the limiting value of $F$, from Eq. (5), is

$$F_\infty = \left[\frac{\beta}{\beta+2}\right]^2 \quad (9)$$

for high frequencies; and

$$F_0 = 1 \quad (10)$$

at low frequencies. The curve is symmetrical if $\beta \gg 2$ and $r \ll Z$, a condition which is usually realized in practice. (The value of $\beta$ is often of the order of 10, $Q_0$ is often several hundred, and $Z/r$ is often 2000 or more.) These limiting cases can be seen from simple reasoning. At zero frequency, the circuit of Fig. 1 becomes that of Fig. 7a. Here it is clear that

$$F = \frac{1}{\left(1+\frac{r}{2Z}\right)^2} \cong 1.$$

For very high frequencies, the circuit of Fig. 1 is similar to that of Fig. 7b, and it is again obvious that

$$F = \left[\frac{\beta}{\beta+2}\right]^2$$

as found above.

If, as in Fig. 5, the inductance is center-tapped to ground, the behavior is slightly different except near resonance. For very high frequencies, the circuit equivalent to Fig. 5 is that of Fig. 7c and

$$F_\infty = \left[\frac{\beta}{\beta+2}\right]^2.$$

For low frequencies, the circuit is that of Fig. 7d and likewise

$$F_0 = \left[\frac{\beta}{\beta+2}\right]^2.$$

Thus there is a slight difference in the low and high frequency behavior of the two cases, i.e., tapped coil or condenser, which disappears when $\beta$ is large. (This difference is more apparent than real, for it arises solely because all the losses have been considered to be connected with the coil and none with the condenser.)

---

[4] The standing-wave ratio is the ratio of the sum of the incident and reflected wave peak amplitudes to their difference.

It is interesting to compare the expression (8) for $F$ with that for the *unbalanced* network ($R$ infinite). Details will be omitted. The result is, with $\beta$ preserving the previously assigned meaning and value,

$$F' = \frac{4[Q_0^2(1-m^2)^2+m^2]}{4Q_0^2(1-m^2)^2+m^2(\beta+2)^2}. \qquad (11)$$

Equation (8) reduces to Eq. (11) in the limit of large $\beta$ as it must, i.e., when the equivalent shunt resistance of the resonant circuit is very large. However, $F'$ approaches unity at both very high and very low frequencies and thus is symmetrical about $m=1$.

The band width of the infinite-rejection, or balanced, filter may be defined as the difference in the two frequencies where $F$ has half its value at infinite frequency. From Eq. (8) this requires that

$$\frac{(1-m^2)^2}{(1-m^2)^2+\left[\dfrac{\beta+2}{2Q_0}\right]^2} = \tfrac{1}{2}. \qquad (12)$$

The term $(1-m^2)$ can be replaced closely by $2\delta m$, where $\delta m$ is the shift of frequency from the ($m=1$)-value. Thus Eq. (12) becomes

$$\delta m = \pm(\beta+2)/4Q_0,$$

and the band width at the half-power, or 3-db-down, point is thus

$$B = \frac{\beta+2}{2Q_0} f_0 \text{ cycles/sec.}, \qquad (13)$$

where $B$ is the band width and $\omega_0 = 2\pi f_0$.

For the unbalanced filter in this notation the band width is given simply by

$$B' = \frac{[(\beta+2)^2-8]^{\frac{1}{2}}}{2Q_0} f_0 \text{ cycles/sec.} \qquad (14)$$

and hence reduces to Eq. (13) for large values of $\beta$. A sketch of the antiresonance behavior of the balanced and unbalanced filter is given in Fig. 8. In this figure the minimum in the unbalanced filter curve is given by Eq. (11) while, of course, for the balanced case,

$$F_{\text{res}} = 0. \qquad (15)$$

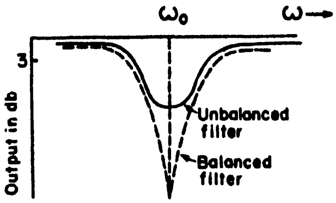


FIG. 8. Schematic behavior of antiresonance curve.

If one defines

$$\gamma \equiv Q_0/Q_L, \qquad (16)$$

where $Q_L$ is the loaded $Q$ of the circuit (because of line loading and external resistance loading) $\gamma$ can be related to $\beta$ by

$$\frac{\beta+2}{2} = \gamma. \qquad (17)$$

The value of the loaded $Q$ of the unbalanced filter at resonance may be seen almost from inspection of Fig. 9. The quantity $Q_0$ is defined as $Q_0 \equiv \beta Z/L\omega_0$; $Q_L$ is likewise defined as $R_1/L\omega_0$, where $R_1$ is the parallel resistance across the terminals of the inductance, i.e., $R_1 = 2\beta Z/(2+\beta)$. Hence:

$$\gamma \equiv Q_0/Q_L = (\beta+2)/2.$$

Now Eq. (11) may be rewritten at resonance as

$$F_{\text{res}}' = \left[\frac{2}{\beta+2}\right]^2 = \frac{1}{\gamma^2}, \qquad (18)$$

which can be easily visualized. The depth of the attenuation cut taken by the filter depends only on the ratio of loaded to unloaded $Q$, and hence principally on the coefficient of coupling of the line to the filter. In principle, infinite attenuation at resonance could be obtained with the unbalanced filter by enormous coupling to the line. However, in such a case, the band width likewise

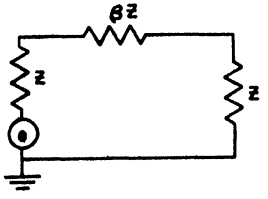


FIG. 9. Circuit for determining loaded $Q$.

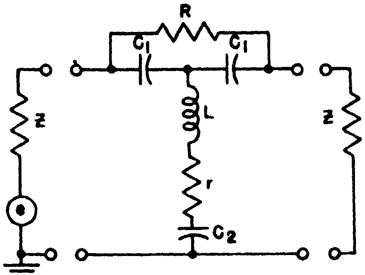


Fig. 10. Circuit equivalent to Fig. 5.

becomes enormous; for Eq. (14) may be written as

$$B' = \frac{f_0}{2Q_0}(\gamma^2 - 2)^{\frac{1}{2}} = \frac{f_0}{2Q_0}\left[\frac{1}{F_{res}'} - 2\right]^{\frac{1}{2}}, \qquad (19)$$

and as $F'$ decreases $B'$ increases.

In the case of the balanced circuit, while $F_{res} = 0$, the band width however is still finite as found above in Eq. (13); *viz.*, in this notation,

$$B = \gamma f_0/2Q_0 = \frac{f_0}{2Q_L}, \qquad (20)$$

and the attenuation at high frequencies, from Eq. (9), is

$$F_\infty = \left[\frac{\gamma - 1}{\gamma}\right]^2. \qquad (21)$$

It is interesting now to put in some figures for a possible u.h.f. application. Let the desired band width be ±0.5 Mc/sec. at 3000 Mc/sec. Thus $Q_L = 3000$. Let $Q_0 = 30{,}000$, and hence $\gamma = 10$. Then, in this case, according to Eq. (21),

$$F_\infty = 0.81,$$

which represents an insertion loss of about 1 db and, of course, $F_{res} = 0$. For the same filter unbalanced,

$$F_{res}' = 0.01,$$

and the band width would be ±0.5 Mc/sec. within 1 percent for both filters.

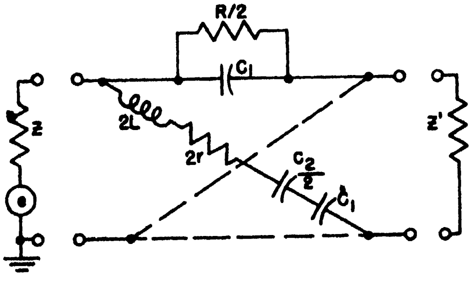


Fig. 11. Lattice development of Fig. 10.

The circuits of Fig. 1 and of Fig. 5, discussed in some detail in this section, by no means exhaust the possibilities of single-section infinite-rejection filters. In fact they are merely prototypes: variants can be obtained by standard transformation theorems. For example, the circuit of Fig. 10 is potentially equivalent to that of Fig. 5. From the symmetrical-lattice development in Fig. 11, it can be shown, with some manipulation, that the condition for infinite rejection is that

$$\omega_0^2 = \omega_2^2 + \frac{r}{R}\omega_1^2, \qquad (22)$$

and

$$R/r = Q_0^2 n^2\left(\frac{n^2}{4} + m^2 - 1\right), \qquad (23)$$

where

$$Q_0 = \frac{L\omega_0}{r}, \quad \omega_2^2 = \frac{1}{LC_2}, \quad \omega_1^2 = \frac{2}{LC},$$

$$n = \frac{\omega_2}{\omega_0}, \quad n = \frac{\omega_1}{\omega_0}.$$

The extension to the case where the capacitances $C_1$ are not equal is readily accomplished but leads to nothing essentially new.[5]

---

[5] Some interesting asymmetry properties are present in this filter. For instance, the possible values of $R/r$ range from 4 to ∞; i.e., when $m = 0$ the value of 4 obtains, and when $m = 1$, $R/r \rightarrow \infty$ and $rn^2/R \rightarrow 0$. In the limiting case of small values of $n^2 (C_1 \gg C_2)$, Eqs. (22) and (23) lead to

$$\omega_0^2 \cong \omega_2^2, \qquad (a)$$

$$R = 4r. \qquad (b)$$

When $n^2$ becomes very large $(C_1 \ll C_2)$, the limiting value of the frequency, again is

$$\omega_0^2 \cong \omega_2^2, \qquad (c)$$

but now we have

$$R = rQ_0^2 n^4/2 \qquad (d)$$

and

$$\omega_1^2/\omega_0^2 \cong (R/r)^{\frac{1}{2}}/Q_0. \qquad (e)$$

(Cf. Eq. (d) and Eq. (2).) From Eqs. (d) and (c) it can be seen that the apparent pull of the resonant frequency from that of the shunt arm alone is somewhat illusory, despite Eq. (22). Thus one possible small advantage ensues. For a given value of $R$ infinite rejection may be obtained by varying $C_1$, and, hence, $n$. If the frequency pulling (which is small as noted) is not serious, this may be a convenient method.

For reference, we write the expression for the fraction of the unfiltered energy transmitted by this balanced filter as a function of frequency.

$$F'' \cong \frac{(f^2-1)^2 R/r}{\left[\frac{R}{4r} + \frac{R^2}{4}\right]\left[\frac{n^2(1+2r)^2}{4Rr} + \frac{4}{n^2}\left(f^2 - m^2 - \frac{n^2}{4}\right)^2\right]},$$

where $f = \omega/\omega_0$.

While the circuit of Fig. 10 may be better than that of Fig. 1 or Fig. 5 for some applications, it suffers from two defects: (1) six circuit elements are necessary *vs.* five in the former case, a defect of some importance in u.h.f. applications; (2) the frequency of infinite rejection only approximates the resonant frequency of the shunt resonant element, and each adjustment of $R$ requires readjustment of $\omega_0$. Furthermore, $R$ cannot be readily visualized or calculated in terms of standing-wave ratios.

Finally, it should be mentioned that, for tuning the balanced filter of Fig. 1 over a comparatively narrow range of frequencies, it is desirable to change both $L$ and $C$, although either one alone can be varied. The condition of balance, from Eq. (2), may be written as

$$R=\frac{\omega_0{}^2L^2}{4r}$$

and, if one eliminates $\omega_0{}^2$ by noting that near resonance $\omega_0L\cong 2/C\omega_0$,

$$R\cong L/2rC. \tag{24}$$

Since $r$ represents the losses in the tuned circuit, it is practically constant during the tuning, and hence the filter remains balanced with $R$ fixed if $L$ and $C$ are changed together in such a way that their ratio stays constant. If, however, $L$ alone is varied or $C$ alone is varied as the frequency is varied, say by $l$ percent or $c$ percent, respectively, for a 1 percent change in $\omega_0$, $R$ must be changed by $(2+l)$ percent or $(1+c)/2$ percent, respectively. Depending on the method and difficulty of changing $R$, we can attempt to tune in one of the three ways.

## III. APPLICATION TO u.h.f. FILTERS

In the previous section we found that a filter of unlimited attenuation could be realized practically, at least for cases where the circuit constants are lumped. The following question now arises: what modifications are necessary when one deals with corresponding distributed-parameter circuits, i.e., u.h.f. lines and u.h.f. resonant cavities? It turns out that this same analysis is applicable, for it is well known that a single-mode cavity resonator can be completely specified in terms of $Q$, $\omega_0$, and $R''$ and, likewise, a u.h.f. line has a definite input impedance (usually real). The cavity resonator can be shunted across the line, or placed in series with one side of the line. In the latter case we have exactly the unbalanced filter of Fig. 1 or Fig. 5 discussed above. The only remaining question is how we can arrange the analogue of $R$, for there is no obvious condenser or inductance to be center-tapped.

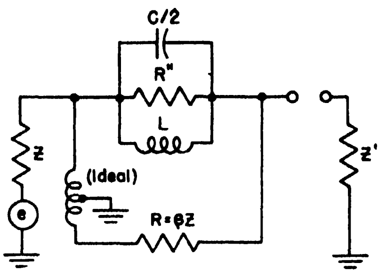


FIG. 12. Infinite-rejection filter in the form of a three-arm bridge.

Before proceeding to a discussion of the methods of realizing the balanced filter it is worth while to introduce still another circuit which can be shown to be precisely equivalent to the balanced filter.

Figure 12 shows essentially a three-arm bridge. Let us assume an ideal 1:1 (auto) transformer such that the impedances are infinite, and close coupled, so that the coefficient of coupling is unity. One is interested in the load current as a function of $R$ and $\omega$. This is given by

$$i_L=\frac{e\beta^2Z^2\left(1-\omega^2\dfrac{LC}{2}\right)}{Z(\beta+2)\left[2\beta Z^2\left(1-\omega^2\dfrac{LC}{2}\right)+j\omega LZ(\beta+2)\right]}, \tag{25}$$

where $i_L$ is the load current and all other symbols have their previous meanings and $R$ is taken

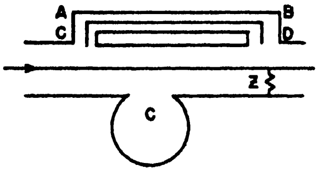


FIG. 13. u.h.f. realization of Fig. 12.

equal to $R''$, i.e., $\beta Z$. Again we form the fraction

$$F=\frac{4|i_L|^2Z^2}{e^2}$$

and find

$$F=\frac{4\beta^2(1-m^2)^2}{(\beta+2)^2\left[4(1-m^2)^2+\frac{(\beta+2)^2}{Q_0^2}m^2\right]}. \quad (26)$$

Remembering that $m^2$ is virtually unity, we note that Eq. (26) is identical with Eq. (8). Thus the attenuation for low frequencies is given by the ratio of $\beta$ to $(\beta+2)$, just as was found before. The band width to the 3-db-down points therefore must be exactly the same. Hence it may be concluded that the bridge circuit is precisely equivalent to the balanced-filter circuit, even to the necessary loss of resonance.

This suggests at once a simple method to obtain the balanced-filter behavior at u.h.f. Consider the circuit of Fig. 13; $C$ is a cavity which is equivalent to the lumped resonant circuit of Fig. 12. The power absorbed in the load $Z$ is made up of two parts: one part comes directly through $C$ to $Z$, the other through the branch line to $Z$. This latter suffers a 180° change of phase by making $\overline{AC}$ and $\overline{BD}$ each a quarter wave-length. Then the power through the branch path arrives at $D$ with phase reversed from that of the power arriving directly. The amount of power bypassed is controlled by the probes in the stub lines. This circuit, as a little thought will show, is the u.h.f. analog of Fig. 12 and hence should possess its advantages. One device of this type has been constructed and will be discussed below.

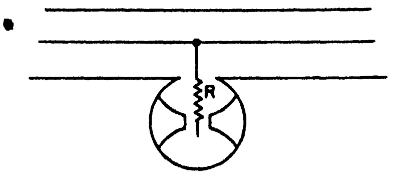


FIG. 14. u.h.f. analog of Fig. 1.

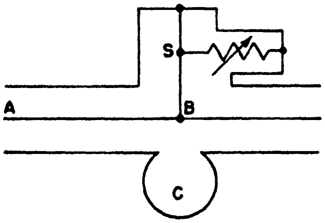


FIG. 15. Second u.h.f. analog of Fig. 1.

Another application of these ideas to a u.h.f. filter is suggested by Fig. 1. Can this circuit be carried directly over to ultra high frequencies? Suppose one had a cavity as in Fig. 14. It is quite clear intuitively that by symmetry if a probe is placed midway between the interior posts of the cavity and connected by a resistor to the center conductor, i.e., ground, it is the complete analog of Fig. 1. However, this procedure is difficult experimentally. Alternatively we might connect $R$ from ground, not to the center point between the posts, but to a point in the coupling window at the midpotential of the window. Thus we would be reasonably sure of obtaining again the analog of Fig. 5. Finally, we may incorporate this idea in a practical way as follows: $C$ again is the cavity (Fig. 15), and $A$ the center conductor of a coaxial line. This is stub-supported at the point $B$, symmetrically located with respect to the (symmetrical) cavity $C$. Now, if the stub center conductor $S$ could be made lossy, with a real resistance between $B$ and ground equal to $rQ_0^2/4$ (where $r$ is the equivalent series loss of the cavity and $Q_0$ is the unloaded $Q$ of the cavity), we would expect the balanced-filter results to apply. This can be achieved by coupling a real load to the stub cavity as shown. By adjusting

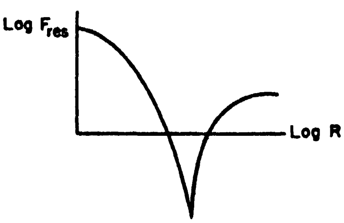
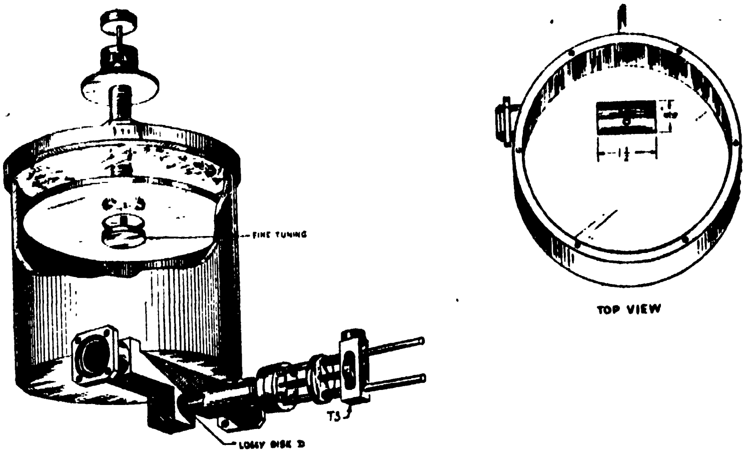


FIG. 16. Log$F_{res}$ vs. log$R$.

FIG. 17. Infinite-rejection filter.



the coupling, choosing the load properly, and placing it at the proper distance from the probe, any equivalent real value of impedance can be introduced into $S$, and the reactive part can effectively be absorbed in the cavity reactance. Thus, since $B$ is essentially at the midpoint of the effective voltage across the cavity $C$, we can balance the filter by this adjustment and by tuning $C$. It should also be noted that there should be no "frequency-pulling" (to the extent that the added reactance at $B$ is small) because the balanced filter and unbalanced filter have the same resonant frequency (see Fig. 8). Furthermore it is not essential that the voltage across the iris be completely symmetrical with respect to $B$. (Note discussion in first part of Section II.) A filter of this design has been constructed, and the data obtained from it will be presented below.[6]

A word should first be said, however, concerning how critical the tuning procedure can be. The sharpness of tuning as $\omega$ is varied can be seen from Fig. 8. In practice, however, this adjustment is usually not very difficult, for one may tune the cavity by stubs moving on fine threads in a region of weak field. The sharpness as $R$ is varied can be calculated. The result is

$$F_{\text{res}}=\frac{4\left(1-\dfrac{4R}{\beta}\right)^2}{\dfrac{16}{Q^2}(R+1)^2+\left[(2R+1)\left(2+\dfrac{4}{\beta}\right)\right]^2} \tag{27}$$

to the same degree of approximation as previously used.

It is noted that $F_{\text{res}}$ becomes zero, as it should, for $R=\beta/4$. A logarithmic plot of $F_{\text{res}}$ is given in Fig. 16. The adjustment of $R$ is thus seen to be critical and, experimentally, needs a little care. (For the curve drawn, $Q$ was assumed 3000 and $\beta=20$; then one can simplify further to obtain

$$F_{\text{res}}\cong\left[\frac{2\left(1-\dfrac{4R}{\beta}\right)}{[2R+1]\left[2+\dfrac{4}{\beta}\right]}\right]^2.)$$

For any particular cavity it is wise to choose the load and load position (or, failing this, to use the proper transformer) such that $R$ at $B$ is approximately the correct value to start with.

In order to retain full benefit of the infinite-rejection circuit, it is clear from Eq. (20) that the loaded $Q$ should be high, and thus for large values of $\beta$, the unloaded $Q$ must be very high. For a band width of 0.5 Mc/sec. at 3000 Mc/sec., a loaded $Q$ of 6000 is needed, and clearly a cavity operating in the $TE_{0,1,1}$ mode[7] is indicated in order to obtain the necessary $Q_0$ of 30,000 or more so as to have a small loss off resonance (see Eq. 21). Conveniently enough, an "echo box" can be modified for this use.[8]

---

[6] Acoustic applications of a filter of this type suggest themselves but will not be discussed here.

[7] See W. L. Barrow and W. W. Mieher, Proc. I.R.E. **28**, 184 (1940).

[8] The "echo box" is merely a high $Q$ cavity. The ratio of stored to dissipated energy is so high that, once excited, the echo box will "ring" for a considerable time.

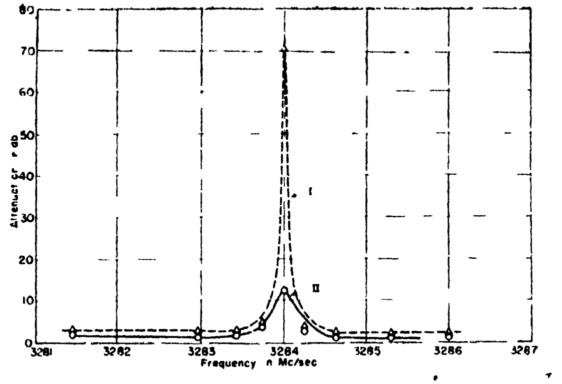


FIG. 18. Filter characteristics. Curve I, balanced filter; Curve II, unbalanced filter.

Although the tuning of the filter is exceedingly sharp for very high rejection ratios (cf. Figs. 8 and 16), and therefore the standard Radiation Laboratory echo box (1943) had to be eventually supplanted by a cavity of similar design but of $\frac{1}{4}$-in. steel plate for the sake of mechanical rigidity, the following results will be reported from a modified standard echo box. Figure 17 is a drawing of the cavity. The loss in the side stub is provided by the resistive disk $D$, whose resistance was adjusted by scraping to the (roughly) calculated value. Further adjustment was provided by the tuning stub $TS$ which, if $R_0$ is the value of the resistance, makes the real impedance at $B$ any value from 0 to $R_0$. Thus the resistance at $A$ can be adjusted to the proper value. (The inevitable reactance introduced can be simply absorbed in that of the cavity.) Iris coupling to the $TE_{0,1,1}$ mode requires the rather peculiar off-set shown. With the filter coupled between the generator and receiver (through a standing-wave detector) by considerable lossy line, curves of filter output *vs.* frequency were obtained. These are shown in Fig. 18. Curve I is for the filter adjusted for infinite rejection; Curve II is for $R$ considerably different from the correct value. It is seen that Curve II is the usual rejection-filter characteristic. Curve I is virtually identical with it at the half-power points and above, but increases the rejection ratio to very large values. The 70-db point was the last measured value and was limited by the receiver sensitivity. There is nothing to make one believe it could not be extended considerably further, although even at the 70-db point considerable difficulty is encountered merely in the frequency instability of the (klystron) signal generator. Table I summarizes the experimental measurements and compares them with the values computed from theory. It is clear that the agreement is satisfactory.

TABLE I. Comparison of theoretical and experimental results.

| Design characteristics | Modified echo box | | Cavity with center posts | | Probe-type filter | |
|---|---|---|---|---|---|---|
| $f_0$ in Mc/sec. | 3284 | | 3160 | | 3233 | |
| $\beta$ | 5.78 | | 16.8 | | 2.7 | |
| $B'$ in Mc/sec. | 0.5 | | 14.7 | | 1.04 | |
| | Exp. | Theor. | Exp. | Theor. | Exp. | Theor. |
| $\gamma$ | 4.2 | 4.63 | 10 | 9 4 | 3.6 | 2.35 |
| $B/B'$ | 1.02 | 1.05 | 1.14 | 1.01 | 1.15 | 1 21 |
| Insertion loss in db | 1.1 | 1.06 | 2 25* | 1 0 | 6.1** | 5 |
| $F_{\text{res}}/F_{\text{res}}'$ in db | 60 | ∞ | 35 | ∞ | 11 | ∞ |

* There was an asymmetry in the resonance curve due to imperfect matching between line and load. This partly accounts for this discrepancy.

** Averaged over asymmetry.

A word might be said of two preliminary constructions. To check the theory, a cavity of the type sketched in Fig. 14 was constructed as a balanced filter. Figure 19 describes it, and Table I gives the experimental and theoretical results obtained with it. For cases where extreme sharpness is not required this filter is easy to make, easy to adjust, and easy to operate.

Finally, the u.h.f. analog of Fig. 12 has likewise been constructed from a modified echo box. This is shown in Fig. 20. Because of inability to couple properly by means of probes, the tabulated theoretical and experimental results shown in Table I are not in as good agreement as in the other two cases.

## IV. PULSE DISTORTION PRODUCED BY FILTER

Although the results of investigations into the distortion introduced in an initially rectangular pulse emerging from a single-tuned filter (pass or rejection type) are probably well known for the case where the filter is tuned to the carrier frequency of the pulse, not as much information seems to be available for the case where the filter is not tuned precisely to the carrier.[3]

The mathematical expression for the transfer impedance of an infinite-rejection filter becomes too involved for tractability in Fourier analysis. Thus one is forced to idealize the actual problem. This can be done, however, with practically no modification of the actual state of affairs, for the attenuation characteristic of an infinite-rejection filter very closely resembles that of a single-tuned circuit with a $Q$ far too high to be realized experimentally. Thus the following is required of the single-tuned circuit: (1) it shall have the same band width as the infinite-rejection filter, and (2) it shall have the same attenuation at resonance. The correspondence in shape then follows automatically. The experimental filter described in Section III had a band width of 0.5 Mc/sec. at 3000 Mc/sec., with a loss off resonance of approximately 1 db. Let us take the attenuation at resonance as 60 db. Then the appropriate constants for the idealized filter are $Q_0 = 6\times 10^6$ and $R = 2000$, where $Q_0$, the unloaded $Q$ of the filter, is defined by $Q_0 = R/L\omega_0$. The term $L$ is the equivalent lumped inductance of the filter, and $R$ is the equivalent shunt resistance, measured in units of line impedance.

The system function, i.e., ratio of output to input voltage, is given by

$$h(\omega) = \frac{1 + 2j\dfrac{Q_0}{\omega_0}(\omega - \omega_0)}{2 + R + 4j\dfrac{Q_0}{\omega_0}(\omega - \omega_0)}, \tag{28}$$

where $j \equiv (-1)^{\frac{1}{2}}$, and $\omega_0$ is defined by

$$\omega_0^2 \equiv \frac{1}{LC}.$$

Now, let $e_1$ be a rectangularly modulated sinusoidal carrier of radian frequency $\omega$, with unit amplitude and pulse duration $\tau$ seconds.

Before we proceed to a mathematical formulation of the resulting output voltage $e_2$, we may describe the process in physical terms. The incoming pulse shock-excites the filter. The filter "rings" for the duration of the pulse at the frequency of the pulse, with a steady-state amplitude dependent on the resonance characteristic of the filter. To this steady-state value is added a transient oscillation decaying exponentially to zero at a rate determined by the $Q$ and $R$ of the filter. At the conclusion of the pulse, the steady oscillation disappears, and all that is left is the decaying term, but now to it is added a second decay term out of phase, initiated by the "shock" of returning the carrier level to zero. If, in addition, the filter is somewhat detuned from the carrier (of frequency $\omega_1$) the decay terms, representing self-resonant transient oscillations, beat with the steady-state oscillations at a frequency equal to the difference between the two. Thus the resultant shape of the pulse is cooperatively determined by (a) the $Q$ and $R$ of the filter, and (b) the frequency difference $\Delta f = (\omega_0 - \omega_1)/2\pi$ cycles/sec.

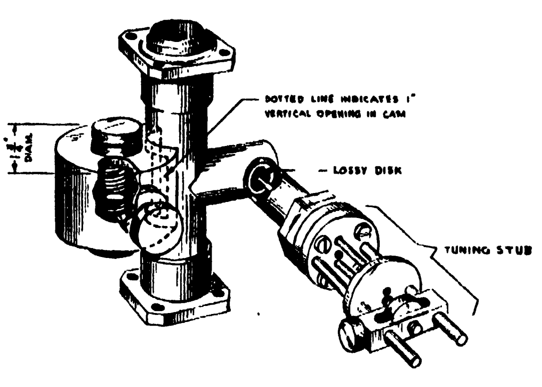


FIG. 19. u.h.f. filter.

Returning to a mathematical formulation of this result, we see that the function $e_2(t)$ can be obtained from $e_1(t)$ by a technique now fairly standard.[9] The Fourier spectral distribution of the unit pulse of duration $\tau$ is multiplied by $h(\omega)$, expanded into partial fractions, and reintegrated with respect to $\omega$ (with, of course, proper attention to the choice of path of integration in the complex plane) In short, there

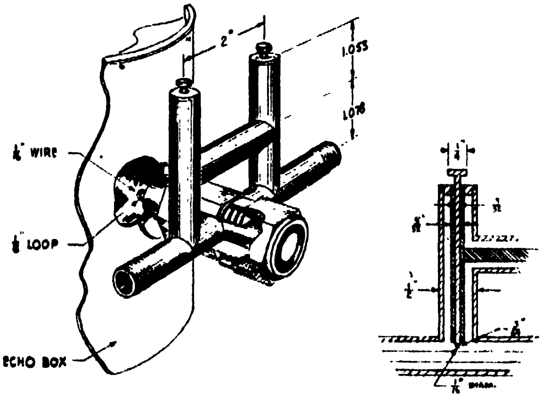


FIG. 20. Bridge-type filter.

---

[9] E. A. Guillemin, *Communication Networks* (John Wiley and Sons, New York, 1935), Vol 2, Chap. XI.

results

$$e_2(t)=\frac{\exp j\omega_1 t}{2+R+4j\frac{Q_0}{\omega_0}(\omega_1-\omega_0)}\left\{1+2j\frac{Q_0}{\omega_0}(\omega_1-\omega_0)+\frac{R}{2}\exp[j(\omega_0-\omega_1)t]\times\exp\left[-\omega_0 t\frac{R+2}{4Q_0}\right]\right\},\quad 0\leq t\leq\tau;\quad (29)$$

$$e_2(t)=-\frac{\frac{R}{2}\exp[j\omega_1 t]}{2+R+4j\frac{Q_0}{\omega_0}(\omega_1-\omega_0)}\left\{\exp[j(\omega_0-\omega_1)t\times\exp\left[-\omega_0 t\frac{R+2}{4Q_0}\right]-\exp\left[-\omega_0\frac{R+2}{4Q_0}(t-\tau)\right]\times\exp[j(\omega_0-\omega_1)(t-\tau)]\right\},\quad \tau\leq t\leq\infty,\quad (30)$$

where, to repeat, $\omega_1$ is the radian frequency of the carrier, $\omega_0$ is the resonant radian frequency of the filter, and the factor multiplying $\exp[j\omega_1 t]$ is the resultant pulse shape.

Expressions (29) and (30) have been calculated for a 1-μsec. pulse and with differences of carrier frequency from filter frequency of 0, ±0.2, ±0.4, ±0.6, ±0.8, ±1, and ±1.5 Mc/sec. It is to be noted, of course, that the result is even in the difference frequency, as would be expected.

The final appearance of the pulse will depend strongly, of course, on further phase shifts, attenuation, and demodulation in the receiver. However, phase shift and attenuation can both be neglected in a well-constructed receiver (except possibly for a linear phase shift which merely shifts the origin of the pulse). A linear detector will have a video output, therefore proportional to $|e_2(t)|$ (disregarding the carrier frequency part). A square-law detector will have an output proportional to $|e_2(t)|^2$, and so on. The curves plotted from Eqs. (29) and (30) and shown in Fig. 21 are calculated from $|e_2(t)|^{\frac{3}{2}}$, i.e., for a three-halves power second detector.

The receiver on which measurements were made was carefully constructed, with an *i-f* band

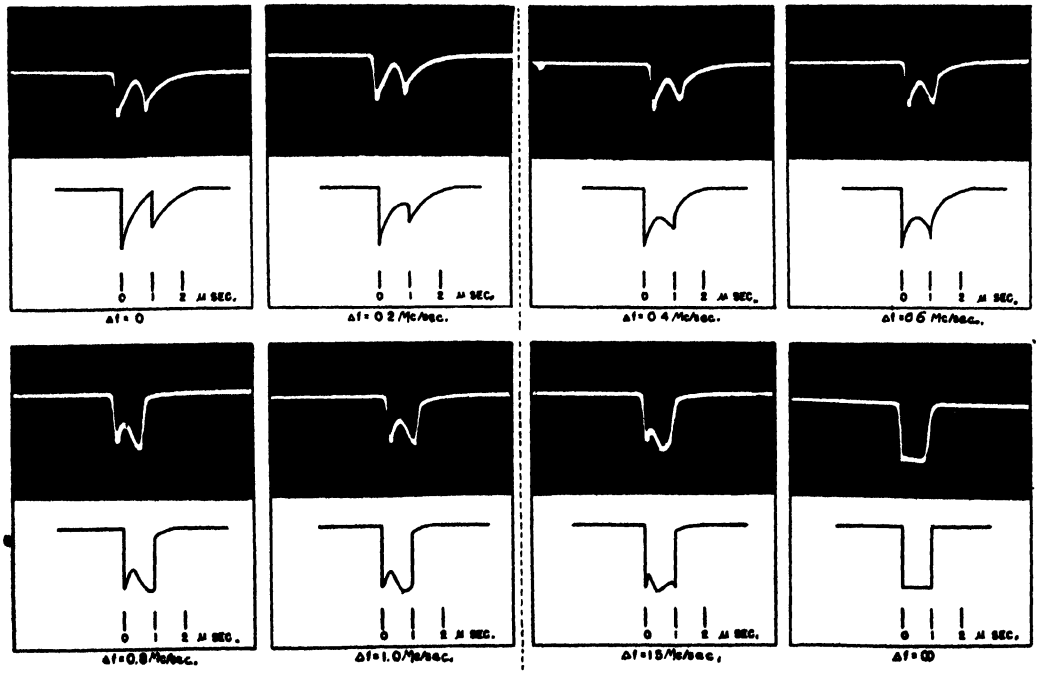


FIG. 21. Distortion produced in rectangular wave by a high $Q$ rejection filter detuned from the pulse frequency by $\Delta f$ Mc/sec.

width of some 13 Mc/sec. and a video band width of some 7 Mc/sec. The second detector was somewhere between linear and three-halves power law. One photograph reproduced in Fig. 21 shows the resultant rectangular wave form. The pulse was formed by driving the grid of a klystron amplifier (excited by a klystron oscillator) from cut-off to a conduction value. This resulted in a pulse virtually free from f-m. The output cavity of the amplifier was broadened in bandwidth by loading it externally. The other photographs carry self-evident explanations.

The agreement between experimentally determined and theoretically predicted curves is seen to be quite satisfactory, with the possible exception of the pulse for $\Delta f = 1.5$ Mc/sec., where $\Delta f \equiv (\omega_1 - \omega_0)/2\pi$. Such differences as remain can be accounted for by the lack of abruptness in the rise and fall of the pulse, somewhat inexact guesses as to $Q_0$ and $R$, and the slight difference in the law of the second detector assumed and that used. However, more accurate determination is hardly warranted.

It is clear that for small frequency differences the pulse is spread out to some two microseconds or so, with a more or less triangular shape, but in every case there is a fairly well defined discontinuity at $t = 1$ μsec. While the attenuation of the *center* frequencies in the pulse spectrum is essentially complete, enough energy remains to give respectable signals. One crude experiment showed a loss of about 2 db in the minimum discernible signal in receiver noise for a filter tuned so that $\Delta f = 0$. As $|\Delta f|$ increases, the depth of the cut in the pulse and the length of the tail decrease more or less uniformly until a case is reached where there is about one wiggle (due to the beat phenomenon mentioned above) across the top of the pulse ($\Delta f \cong 1$ Mc/sec.), then two wiggles with reduced amplitude, and so on. The explanation of the discrepancy between the theoretical and experimental curves for $\Delta f = 1.5$ Mc/sec. is probably that the difference frequency differs slightly from the nominal value. The larger $\Delta f$, the more phase change in the wiggles for a given error in frequency setting. The beats finally die out and a square pulse is regained when the filter is tuned far off resonance. It is therefore not hard to determine the tuning of the filter to a fairly considerable accuracy from the video appearance of a square pulse.

# Theory of Automatic Control Systems*

M. AVRAMY MELVIN
*Columbia University, New York, New York*

(Received June 7, 1946)

The theory of *robots* or reproducing power-amplifying systems is considered. After a general terminological orientation the study in the first part is confined to *simple robots*, the theory of which is equivalent to that of nonlinear springs with inertia and friction. The alacrity and fidelity of control are well exhibited by studying the behavior of a robot under constant tracking. Special emphasis is placed on the specification of conditions for *optimum performance*, i.e., minimum deviation or maximum fidelity after any given time. Curves are given for both proportional and on-off robots, from which the parameter values for optimum performance may be read.

In Part II the theory of proportional robots is analyzed in detail, especially from the point of view of conditions for optimum performance. Two subsequent papers will contain the theory for the on-off case and the general nonlinear robot case. The paper is organized as follows:

## PART I. DESCRIPTIVE ACCOUNT

### A. Terminology and Kinematic Description

A POWER ROBOT may be defined as a *reproducing power-amplifying system.* In other words the name refers to any automatic control system which reproduces the time variation of some quantity which is fed into it, but *with enormously amplified power.* Now, power (whether mechanical or electrical, see Glossary—Table I) is measured by a force-like quantity,** $m$, multiplied by the time variation of some position-like quantity, $\theta$, i.e.,

$$\text{Power} = m\frac{d\theta}{dt} \equiv mv.$$

The robot may then be said to reproduce the time variation, $v$, of the position of some guide, $\theta_g$, with an enormously increased $m$ so that a heavy *load* may be made to follow. The manner in which this is accomplished is by having a delicate *pick-up* operate a *motor* which applies power to the load. The pick-up is so constructed as to compare the position, $\theta_g$, of the guide with $\theta$ of the load and apply a restoring force $m'$ by

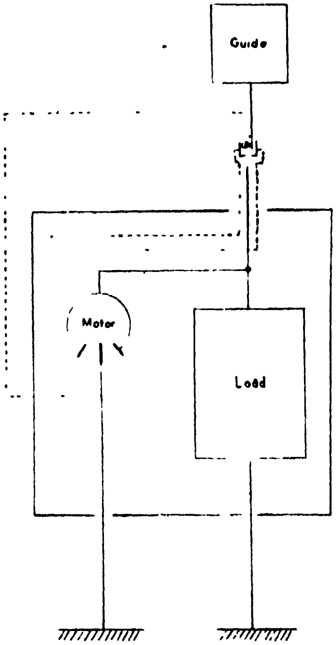


FIG. 1. Schematic diagram of a simple robot.

---

* This work was done in 1943–44 under Contract No. OEMsr-618 between Columbia University and the Office of Scientific Research and Development which assumes no responsibility for the accuracy of the statements contained herein. It was carried out with practically no reference to existing literature, which was, in any case, largely unavailable. Any deviations of terminology from that current among the workers in the field may be ascribed to this. It is hoped that the extensive glossary at the beginning will compensate somewhat for this deficiency. The terms used were carefully chosen with a view to suggestiveness, and some may recommend themselves to others.

** Apologies are here made for the notation, originally adopted in unpublished preceding reports for other reasons. It is to be noted that the theory is much more general than appears here, since $m$ and $\theta$ may be any pair of conjugate physical quantities (original report). But, for simplicity, one may think of $\theta$ as a position variable (or angle) and of $m$ as a force (or torque). The system then reproduces the motion or "tracking" of a *guide*, which is detected by the *pick-up* and applied *via* a *motor* to the *load* or controlled system.

means of the motor to reduce the magnitude of the difference, $D'=\theta_g-\theta$, between them. We shall call this difference the ***output deviation***, and shall designate as a ***simple robot*** any system in which the operation of the pick-up is made to depend directly on this output deviation, $D'$. Other more general types, such as *transmission* robots, may be discussed in another paper.

A schematic diagram of a simple robot is given in Fig. 1. It will generally be arranged that all reaction of the controlled system back on the guide will be negligibly small. We shall further include in the specification of a simple robot the physically natural requirement:

***m′ is a single valued integrable function of D′ in any finite domain.*** Let the phrase ***apparent work*** (i.e., as measured in the frame of reference moving with $v_g$) describe the integral $\int m'dD'$. We may then formulate the condition in words: ***The apparent work done against the motor is finite in any finite deviation and is zero upon the completion of any closed cycle.***

In general one may write the equation of motion of a simple robot as

$$Z[\theta]=m'(D'),$$

where $Z$ is what we may call the "displacement operator" or "stiffness" of the load (usually involving differential operators) and $m'$ is the restoring force function. In very many cases, including all those considered in this paper, the displacement operator is linear in $\theta$ and its derivatives (or integrals). Examples are translational inertia, rotational inertia about an axis, "1st power law" viscous resistance, magnetic damping, Hooke's law elastic stiffness, etc. In all such cases we may rewrite the equation of

TABLE I. Glossary and examples of terms used.

| Symbol | Description | Translational motion | Rotational motion | Electric circuits | Dimensionless forms for usual load case |
|---|---|---|---|---|---|
| $T$ | total time | | | | $\mathbf{T}=BT/I$,<br>$\mathcal{T}=m_0T/v_gI$ |
| $t$ | time from beginning of a cycle | | | | $\tau$ |
| $\theta$ | position-like variable | position | angle | work per unit current or time integrated voltage | $\phi=B^2\theta/m_0I$ |
| $v=d\theta/dt$ | velocity-like variable | velocity | angular velocity | voltage | $\omega=Bv/m_0$ |
| $m$ | force-like variable | force | torque | current | $\mu=m/m_0$ |
| $I$ | inertance | mass | moment of inertia | capacitance | |
| $B=1/R'$ | friction coefficient $=\dfrac{1}{\text{respondance}}$ | resisting force per unit velocity | torque per unit angular velocity | conductance $=\dfrac{1}{\text{resistance}}$ | |
| $K=1/L'$ | stiffness $=\dfrac{1}{\text{compliance}}$ | elastic force per unit displacement | elastic torque per unit angle turned | inverse inductance | |
| $g$ | as subscript indicates quantity associated with guide | | | | |
| $b$ | as subscript indicates quantity at beginning | | | | |
| $n$ | as subscript indicates quantity associated with $n$-th cycle if there is oscillation | | | | |
| $n+\frac{1}{2}$ | as subscript indicates quantity associated with second half of $n$-th cycle if there is oscillation | | | | |
| $\infty$ | as subscript indicates quantity associated with steady state | | | | |
| $D'=\theta_g-\theta$ | total output deviation | | | | $\Delta'=\dfrac{B^2}{m_0I}D$ |
| $D_i=\theta_g-\theta_i$ | input deviation | | | | $\Delta_i$ |
| $D_\infty'$ | mean steady state deviation from guide position | | | | $\Delta_\infty'$ |
| $D=D'-D_\infty'$ | output deviation from steady state deviation = "deviation" | | | | $\Delta$ |
| $\bar{D}_n$ | maximum amplitude of deviation in $n$-th cycle | | | | $\bar{\Delta}_n$ |
| $V=v_g-v$ | time rate of change of deviation | | | | $\Omega=\dfrac{B}{m_0}V$ |
| $T'$ | time measured from the beginning of constant tracking | | | | $\mathbf{T}'$ |
| $T_0'$ | interval between beginning of constant tracking and first vanishing of $D$ | | | | $\mathbf{T}_0'$ |
| $T=T'-T_0'$ | time measured from first vanishing of $D$ | | | | $\mathbf{T}$ |
| $P_n$ | period of $n$-th cycle | | | | $\pi_n$ |
| $n(\bar{D})$ | Number-Amplitude function giving number of cycles required for maximum amplitude to fall to $\bar{D}$ | | | | $n(\bar{\Delta})$ |
| $\bar{D}(T)$ | Amplitude-Time function giving maximum amplitude as a function of time | | | | $\bar{\Delta}(\mathbf{T})$ |
| $f=\dfrac{1}{P_n}\left(\dfrac{1}{\lvert V_n\rvert}+\dfrac{1}{\lvert V_{n+\frac{1}{2}}\rvert}\right)$ | Fidelity Fraction: approximate fractional time per cycle spent within unit $D$ interval about $D_\infty'$ | | | | $f_\Delta=\dfrac{m_0I}{B^2}f$ |
| $B_{\min}$ | optimum value of the friction coefficient for minimizing the deviation amplitude at any time | | | | |
| $B_{\max}$ | optimum value of the friction coefficient for maximizing the fidelity fraction at any time | | | | |

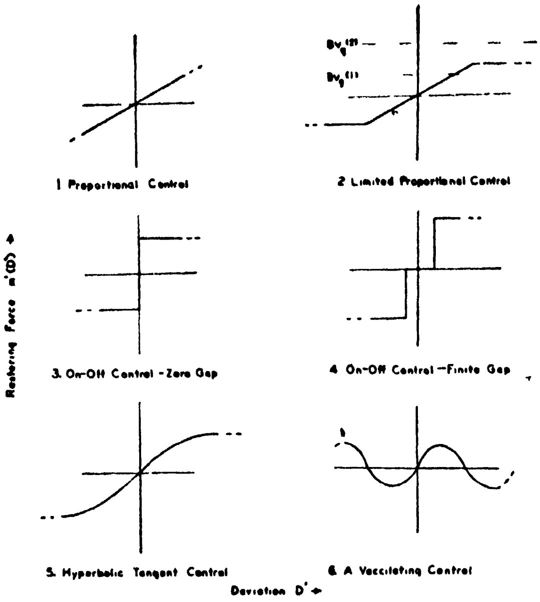


FIG. 2. Restoring force deviation graphs for various types of control.

motion in terms of $D'$:

$$Z[D'] = z(t) - m'(D')$$

where (1.01)

$$z(t) = Z[\theta_g(t)]$$

is the result of operating with $Z$ on the guide position as a function of time.

In general the pick-up is such that the restoring force (i.e., "cause") is not linearly proportional to the deviation ("effect"), so that the theory of robots will be in the domain of non-linear dynamics. The special linear case in which the pick-up does act to produce a restoring force proportional to the deviation is designated by the name of ***proportional control.*** It will be analyzed below so that it may serve as a standard of comparison for other, non-linear controls. The graphs of restoring force *versus* deviation for several types of control are given in Fig. 2.

A division which should be made from the point of view of a general theory is between ***faithful*** and ***vacillating*** (or faithless) robots as illustrated in Fig. 2. In a later paper[1] we shall give a precise and general definition of faithfulness, but tentatively we may indicate the distinction as follows: In a faithful robot the restoring force always acts so as to reduce the deviation, i.e., $m$ is always positive for positive $D$. Not so for a vacillating robot, a type which may occur, for example, in rotating control systems where, if the deviation exceeds a certain value, a new guide (or the old one from the other side) takes over as the center of attraction and the old one is deserted. Certain theorems which may be proved, such as the approach in constant tracking of every simple robot to a steady state of zero oscillation, may be proved only for faithful robots. The others may vacillate throughout eternity.

In any case, the efficiency of a robot will be measured by the ***fidelity with which the load is made to follow the guide,*** i.e., by the smallness of the ***time lag*** and ***oscillation*** of the output deviation $D'$ for changes in $\theta_g$. The more ideal the control system, the smaller will these quantities be. It is one of our prime objectives not only to deduce the general description of the phenomena but also to determine what values of the parameters will minimize lag and oscillation.

There are two interesting special cases:

I. The amplification of power associated with an oscillatory motion.

II. The amplification of power in the case of constant tracking.

The first is the case in which the guide moves with a given oscillatory motion, and the problem is to determine what the amplitude, frequency, and phase of the output deviation are. This is a generalization of the problem of forced oscillations in linear mechanics and we shall not discuss it in this paper.

The second case is that in which, beginning at a certain time when the velocity of the load is $v_b$ and the output deviation is $D_b$, the tracking velocity, $v_g$, of the guide becomes and remains constant. It is a generalization of the problem of free oscillations, or oscillations under a constant force, in linear mechanics. It is this ***case of constant tracking*** with which we shall primarily be concerned, for the fidelity of control of a robot is perhaps best brought out by studying its characteristics for this case (Fig. 3). After a certain time and number of cycles (finite or infinite) the system may attain a ***steady state.*** The steady state may be either one of no oscillation, i.e., perfect

---

[1] "Theory of damped non-linear oscillators with applications," referred to in the following as D.N.O.

following of the tracking motion, or of a steady oscillation with finite period and amplitude. The following or oscillation may be about the correct tracking motion or about a position lagging by a certain steady value behind the correct position. When it occurs, we shall denote this mean steady-state deviation from the guide position by the symbol $D_\infty'$ (an explicit definition will be given later), and it will be convenient to separate it out from the total output deviation, so that we may write

$$D \equiv D' - D_\infty'.$$

Here then, $D$ represents the deviation from the mean steady-state tracking position (Fig. 3). Since it is often relatively easy to correct for a steady space deviation like $D_\infty'$, it is desirable to concentrate merely on minimizing the oscillation and time lag of $D$.

When there is oscillation it is often convenient to start counting time and number of cycles from the first time when $D=0$. Let the interval between the beginning of constant tracking and this first zero of $D$ be represented by the symbol $T_0'$ (Fig. 3) so that, if $T'$ represents the time as measured from the beginning of constant tracking, we have

$$T = T' - T_0'.$$

When the general theory of a system for an arbitrary beginning time is worked out, $T_0'$ may be determined by an auxiliary calculation. The advantage of this shift of the origin of time to the beginning of the first cycle is that it will often simplify the ensuing formulas by eliminating a phase-shift term. Sometimes $T_0'$ is of interest for itself, in that a certain event for which the control is motivated in the first place is made to occur at that moment. Even in cases where steady permanent control is desired, it might be that the system could be brought immediately to the steady state at that moment by a suitable impulsive process.

If $D$ does oscillate in time we may wish to analyze its variation in each cycle independently. In such circumstances it will be convenient to start the time scale anew at the beginning of each cycle. The time so measured within each cycle will be denoted by the symbol $t$.

Let us designate the maximum and minimum $D$ in the $n$-th cycle (the *deviation amplitudes*) by $\bar{D}_n$ and $D_n$, respectively. Also let the period in the $n$-th cycle be represented by $P_n$. Upon inverting the expression for $\bar{D}_n$ as a function of $n$, one may arrive at the

*number-amplitude* function: $n(\bar{D})$

giving the number of cycles required for the maximum deviation to fall and remain below $\bar{D}$. The control characteristics of a system are per-

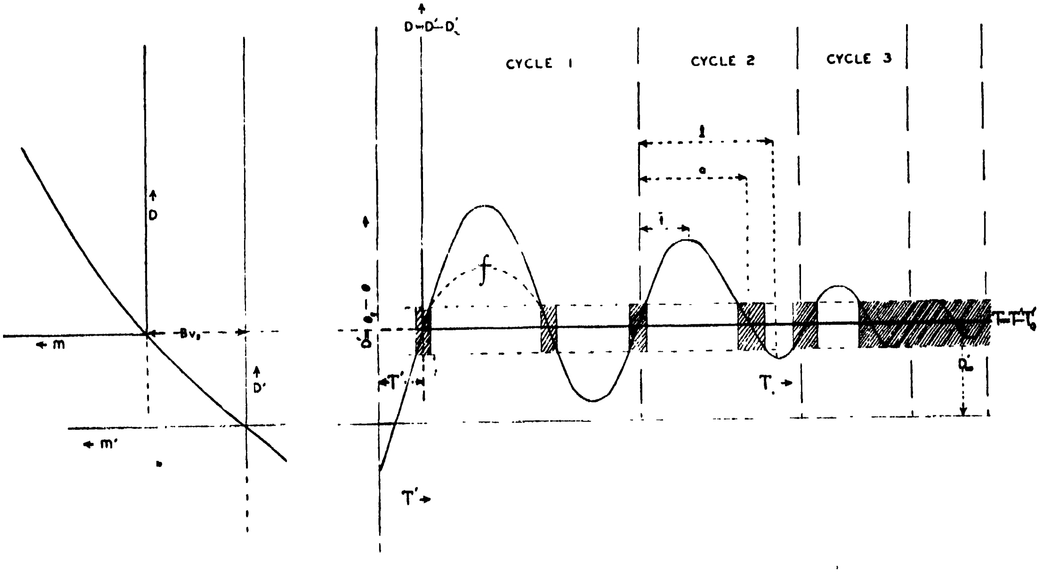


Fig. 3. Oscillations in a simple faithful robot under constant tracking.

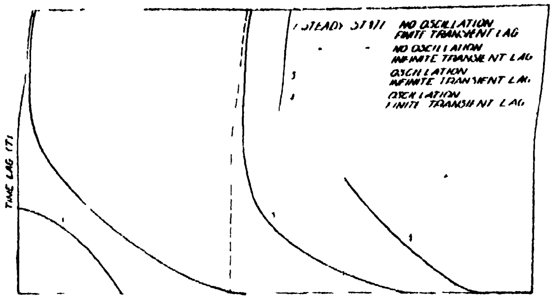


FIG. 4. Time lag *vs.* maximum deviation amplitude.

haps best exhibited by its

$$\textit{time-amplitude curves:}\quad T(\bar{D}) = \sum_{n=0}^{n(\bar{D})-1} P_n,$$

giving the time required up to the beginning of the cycle in which the maximum aplitude falls below $\bar{D}$. These may be plotted for various values of the relevant parameters.

The magnitude $\bar{D}_\infty$, after a steady state has been reached, will be finite or zero, depending on whether there is or is not a steady state of oscillation. The number of cycles and time required to reach the steady state exactly, which we designate

number of transient cycles: $n_\infty$

and

$$\text{transient lag,}\quad T_\infty = \sum_{n=0}^{n_\infty - 1} P_n,$$

respectively, may be finite or infinite. Combining this with the possibility of zero or finite oscillation, we have the four main types of time-amplitude curves one may expect to encounter, sketched in Fig. 4. The functions should actually be represented as discrete, since the variables $\bar{D}$ and $T$ take on discrete values but, for convenience, we have drawn smooth curves through the points.

Besides the $\bar{D}-T$ curves, which set upper limits to the deviations after any particular time, we may be concerned to know what fraction of the time of any cycle the system spends with a deviation in the interval $dD$ between $D$ and $D+dD$. In other words, we should like to know the "distribution function" for the probability of finding the system in the neighborhood of any given deviation about the mean tracking position. Since it is proportional to the time spent within unit interval about $D$, this distribution function is inversely proportional to the time rate of change of $D$ at that point. Thus letting

$$V \equiv \partial D/\partial T,$$

and expressing it as a function $V(D)$ of $D$, we may write:

$$f(D)dD = \frac{1}{P}\left[\frac{1}{|V(D)|_R} + \frac{1}{|V(D)|_L}\right]dD,$$

where we have added together the time intervals for $dD$ when the deviation is increasing and decreasing, respectively, since both contribute to the probability of finding the system in $dD$. division by the period $P$ is merely to insure normalization, i.e.,

$$\int_{-D}^{D} f(D)dD = 1.$$

Because of the nature of oscillation (velocity zero at the amplitude limits) $f(D)$ will be of a general U-shape as sketched in Fig. 5. Now it is difficult to characterize this curve in a general way. For practical purposes, however, it will be of primary interest to know what fraction of the time of any cycle the system spends within any small interval in the ***neighborhood*** of the mean steady-state position, i.e., we wish to give ***the value of the ordinate, f, of the distribution curve at the abscissa,*** $D=0$. This we shall designate the ***fidelity fraction,*** since multiplying it by any small deviation interval $\Delta D$ about $D=0$ gives an approximation to the fractional time spent within that interval. Since the area under the total curve has been normalized to unity, it is clear that $f$ is also represented by the shaded area in Fig. 5. An explicit formula for $f$ in terms of the velocity $V_n$ at the beginning of the $n$-th cycle and the velocity $V$ at the time $a$ of the return passage through zero $D$ is

$$f = \frac{1}{P_n}\left(\frac{1}{|V_n|} + \frac{1}{|V(a)|}\right)$$ [***]

---

[***] As pointed out by the referee, the quantity $f$ here defined is more applicable to the first half of the cycle rather than to the total cycle. It is still, however, an approximate index of fidelity and may be taken as having certain advantages of simplicity. In any case, it had already been used in the treatment of the special cases given later.

## B. Dynamical Equation of Robots with Usual Load

Thus far our considerations have been quite general. But it is convenient here to limit ourselves to the very important common case where the sluggishness of the load or controlled system is governed mainly by inertia and frictional resistance so that the equation of motion of the robot may be written

$$\begin{gathered} I\frac{dv}{dT'}+Bv=m'(D') \\ \left.\begin{aligned} v&=v_b \\ D'&=\theta_u-\theta_b \end{aligned}\right\}\text{at } T'=0, \end{gathered} \tag{1.02}$$

where $I$ is the inertia of the load and the $B$ its frictional resistance per unit velocity and, as is the nature of the robot, the applied force $m'$ is a function of the deviation $D'$. (In the case of a transmission robot, $D'$ must be taken as representing the input deviation.) The subscript $b$ in the initial condition equations represents the values of the indicated quantities at the beginning $T'=0$. In terms of $\theta$, the differential equation may be written

$$I\frac{d^2\theta}{dT^2}+B\frac{d\theta}{dT}=m'(\theta_u-\theta)=m_0\mu'(\theta_u-\theta)$$

where we have taken out a constant *characteristic strength* factor $m_0$ (which may be defined conveniently in any particular type of system) and have left the functional dependence in the dimensionless restoring force function $\mu'(\theta_u-\theta)$.

Now, in developing the theory for this usual load case, it is convenient to choose new dimensionless measures of time, velocity, and force such that

$$\frac{m'}{m_0}\equiv\mu' \quad \frac{B}{I}T'\equiv\mathrm{T} \quad \frac{B}{m_0}v\equiv\omega \quad \frac{B^2}{m_0I}\theta\equiv\phi.$$

Therefore,

$$\frac{B}{I}t\equiv\tau \quad \frac{B}{m_0}V\equiv\dot{\omega}_u-\omega\equiv\Omega \quad \frac{B^2}{m_0I}D'\equiv\Delta' \tag{1.02'}$$

$$\frac{I}{m_0}\cdot\frac{dv}{dt}\equiv\frac{d\omega}{d\tau} \quad \text{etc.,}$$

where

$$\kappa\equiv m_0I/B^2.$$

Thus we are measuring *all forces in fractions of the characteristic strength* and *all times in fractions of the time constant* of the load when slowing down freely under no applied load; and we are measuring *all velocities in fractions of the characteristic steady-state velocity of the load,* i.e., its motion against friction when the constant characteristic strength of the motor is applied steadily. With these notations the basic robot equation (1.02) takes the form

$$\frac{d^2\phi}{d\mathrm{T}'^2}+\frac{d\phi}{d\mathrm{T}'}=\mu'[\kappa(\phi_u-\phi)],$$

or if we express the equation in terms of

$$\Delta'=\phi_u-\phi$$

we get

$$\frac{d^2\Delta'}{d\mathrm{T}'^2}+\frac{d\Delta'}{d\mathrm{T}'}=\left(\frac{d^2}{d\mathrm{T}'^2}+\frac{d}{d\mathrm{T}'}\right)\phi_u-\mu'(\kappa\Delta').$$

This is the form which the fundamental operator equation (1.01) assumes in usual load systems.

If we confine ourselves to the case of constant tracking beginning at $\mathrm{T}'=0$,

$$\phi_u=\omega_u\mathrm{T}'$$

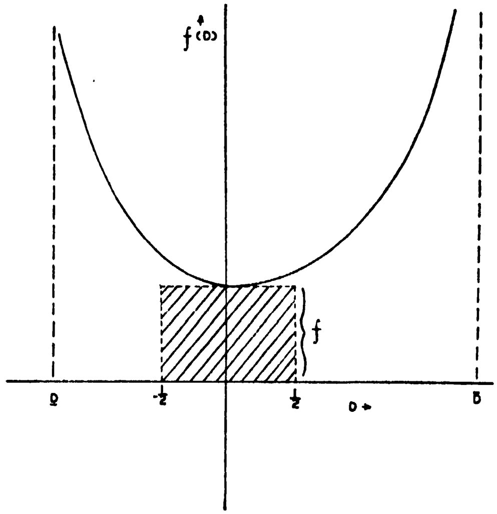
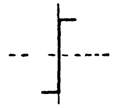
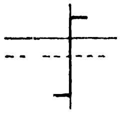
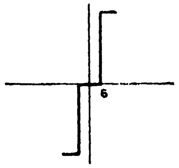
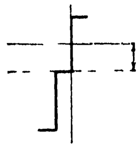


FIG. 5. Distribution function for probability of finding deviation in any neighborhood.

TABLE II. Steady-state deviations and modified force functions for various types of robots.

| Type of control | Restoring force deviation relation $\mu'(\Delta')$ | Steady state deviation $\Delta_\infty'$ | Modified restoring force function |
|---|---|---|---|
| Proportional | $\kappa\Delta'$ | $\frac{\omega_g}{\kappa}$ | $\kappa\Delta$ |
| Quadratic | $\hat{\Delta}'(\kappa\Delta')^2$ | $\hat{\omega}_g\frac{1}{\kappa}(\|\omega_g\|)^{\frac{1}{2}}$ | $\pm\|\kappa\Delta\pm(\|\omega_g\|)^{\frac{1}{2}}\|^2-\omega_g \quad (\pm=\hat{\omega}_g)$ |
| Cubic | $(\kappa\Delta')^3$ | $\frac{1}{\kappa}\omega_g^{\frac{1}{3}}$ | $(\kappa\Delta+\omega_g^{\frac{1}{3}})^3-\omega_g$ |
| *m*-th power | $\hat{\Delta}'\|\kappa\Delta'\|^m$ | $\hat{\omega}_g\frac{1}{\kappa}\|\omega_g\|^{1/m}$ | $\pm\|\kappa\Delta\pm\|\omega_g\|^{1/m}\|^m-\omega_g \quad (\pm=\hat{\omega}_g)$ |
| Exponential | $\hat{\Delta}'[e^{\kappa\|\Delta'\|}-1]$ | $\hat{\omega}_g\frac{1}{\kappa}\ln[1+\|\omega_g\|]$ | $\pm[e^{\|\kappa\Delta\pm\ln(1+\|\omega_g\|)\|}-(1+\|\omega_g\|)] \quad (\pm=\hat{\omega}_g)$ |
| Hyperbolic sine | $\sinh(\kappa\Delta')$ | $\frac{1}{\kappa}\sinh^{-1}\omega_g$ | $\sinh(\kappa\Delta+\sinh^{-1}\omega_g)-\omega_g$ |
|  |  | $\|\omega_g\|<1$: $\quad\|\omega_g\|>1$: |  |
| Hyperbolic tangent | $\tanh(\kappa\Delta')$ | $\frac{1}{\kappa}\tanh^{-1}\omega_g$ or $\pm\infty$ | $\tanh(\kappa\Delta+\tanh^{-1}\omega_g)-\omega_g$ |
| On-off zero-gap |  | 0 or $\pm\infty$ |  $\omega_g \quad (\pm=\hat{\omega}_g)$ |
| On-off finite-gap |  $\delta$ | $\pm\delta$ or $\pm\infty$ |  $\omega_g \quad (\pm=\hat{\omega}_g)$ |

$\hat{a}$ means "sign of $a$" or $\frac{a}{|a|}$

and we have:

$$\frac{d^2\Delta'}{dT'^2}+\frac{d\Delta'}{dT'}=\omega_g-\mu'(\kappa\Delta')\equiv-\mu$$

$$(\mu')\qquad\left.\begin{aligned}\Delta'&=-\phi_b\\ \frac{d\Delta'}{dT'}&=\omega_g-\omega_b\equiv\Omega_b\end{aligned}\right|\text{at } T'=0, \tag{1.03}$$

where $\kappa$, $\omega_g$, $\phi_b$, $\omega_b$, $\Omega_b$ are then constants calculable in terms of the given data, and where we have introduced the symbol

$$\mu\equiv\mu'-\omega_g$$

for the ***modified force function***, i.e., the restoring force measured from the level $\omega_g$ instead of zero.

Now it is demonstrated in D.N.O.[1] that the final steady state in constant tracking of any faithful robot is one of zero oscillation, and consists of a constant steady deviation given by setting the right-hand side of the differential equation (1.03) equal to zero:

$$\omega_g-\mu'(\kappa\Delta')=0$$

or

$$\Delta_\infty'=\frac{1}{\kappa}\mu'^{-1}(\omega_g)$$

where $\mu'^{-1}$ is the inverse function corresponding to $\mu'$. Applications of this result to various types of control are given in Table II. It is thus possible to simplify the system of equations of motion (1.03) by introducing the modified force function and expressing everything in terms of the modified deviation $\Delta$:

$$\mu(\kappa\Delta)\equiv\mu'[\kappa\Delta']-\omega_g=\mu'[\kappa(\Delta+\Delta_\infty')]-\omega_g.$$

This shift of reference level of $\mu$ is illustrated in

Fig. 3. If $\Delta$ becomes zero at some finite time, $T_0'$, before the steady state is reached, i.e., in case of at least one oscillation, it is possible to simplify the equations a little further. Introducing the modified force function, and measuring the time in terms of T from the first zero of $\Delta$, we have:

$$\frac{d^2\Delta}{dT^2}+\frac{d\Delta}{dT}=-\mu'(\kappa\Delta)$$

$$\left.\begin{aligned}&\Delta=0\\&\frac{d\Delta}{dT}=\omega_g-\omega_0\equiv\Omega_0\end{aligned}\right\}\text{at } T=0, \tag{I.04}$$

where, if the solution to the problem for arbitrary beginning data is known, $\omega_0$ or $\Omega_0$ may be calculated readily. Conversely, if one has the solution for general values of $\Omega_0$ as a parameter, one may determine the particular value required to fit the initial data

$$\Delta=-\phi_b-\Delta_\infty' \quad \Omega=\Omega_b \text{ at } T=-T_0'.$$

In general, if $\mu'$ is of any other functional form but linear, the differential equation is non-linear in that the right-hand side is a function of the *dependent* variable, and the solution of the system will require special devices in various cases. We shall discuss this question in D.N.O.[1] and in a subsequent paper[2], but for the present we confine ourselves to noting that, once the problem is solved, all the control characterizing concepts defined above may be found, translated, of course, in dimensionless terms. These are listed for reference in Table III. All these quantities generally depend, of course, on the parameters $\kappa$, $\omega_0$, and $\omega_g$ as well as the indicated arguments.

The visualization of the motion is facilitated if it is represented on a diagram on which $\Omega$ is plotted as ordinate against $\Delta$ as abscissa. This gives the "path curve" in the "state plane" (or "phase space") of our damped non-linear oscillator. The differential equation governing the path of the representative point in the state plane follows from a simple rearrangement of (I.04)

$$\frac{d\Omega}{d\Delta}=-1-\frac{\mu(\kappa\Delta)}{\Omega} \tag{I.04'}$$

with the same initial conditions. The time is now merely a parameter tracing the representative point as it moves along the path curve.

It will be shown in D.N.O.[1] that for every faithful robot the path curve is a continuous right-handed (clockwise) spiral closing in on the point $\Delta_\infty'$ which lies on the $\Delta$ axis. This spiral may be bounded between two comparison spirals beginning at the same initial point and made up of successive pieces of linear oscillator spirals with certain frequencies which may be worked out in each case. In this way follow immediately criteria for oscillation and overdamping, and the possible existence of any finite number of oscillations in the transient state of non-linear oscillators. Upper and lower bounds to the number of oscillations and times required for the deviation to fall to any desired value may then be set.

## C. Optimum Performance Curves

If after solving the constant tracking case as represented in (I.04) for any particular system, and finding for instance

$$\bar{\Delta}(T,\kappa,\Omega_0,\omega_g)$$

and

$$f_\Delta(T,\kappa,\Omega_0,\omega_g),$$

we return to the ordinary measures of time and

TABLE III. Control characterizing concepts translated in dimensionless terms.

| | |
|---|---|
| $\Delta_\infty'$ | mean steady-state deviation from guide position |
| $\Delta$ | deviation from mean steady-state deviation |
| $T_0'$ | time when $\Delta$ first becomes zero after the beginning of constant tracking |
| $\bar{\Delta}_n$ | maximum value of $\Delta$ in $n$-th cycle |
| $\underline{\Delta}_n$ | minimum value of $\Delta$ in $n$-th cycle |
| $\pi_n$ | period of oscillation in $n$-th cycle |
| $n(\bar{\Delta})$ | number-amplitude function |
| $\bar{\Delta}(T)$ | amplitude-time function |
| $n_\infty$ | number of transient cycles required to reach the steady state |
| $T_\infty$ | transient lag, i.e., time required to reach the steady state |
| $f_\Delta=\frac{1}{\pi_n}\left[\frac{1}{\lvert\Omega_n\rvert}+\frac{1}{\lvert\Omega(a)\rvert}\right]$<br>$\equiv\kappa f$ | fidelity fraction, i.e., approximate fractional time per cycle spent within unit $\Delta$ interval about steady-state deviation |

---

[2] "Theory of on-off and relay control systems," J. App. Phys. (To be published.)

position, we may write

$$(D)\quad \bar{D}=\frac{m_0I}{B^2}\Delta\left(\frac{B}{I}T,\ \frac{m_0I}{B^2},\ \frac{B}{m_0}V_0,\ \frac{B}{m_0}v_g\right),\qquad (1.05)$$

$$(f)\quad f=\frac{B^2}{m_0I}f_\Delta\left(\frac{B}{I}T,\ \frac{m_0I}{B^2},\ \frac{B}{m_0}V_0,\ \frac{B}{m_0}v_g\right).\qquad (1.06)$$

It is interesting to investigate the dependence of $\bar{D}$ upon the various system parameters. We shall illustrate by considering the dependence upon the friction coefficient $B$, whose influence on the times required for reducing $\bar{D}$ and for increasing $f$ may be considerable. It is then important practically to determine (with a given time) for what value of $B$, $\bar{D}$ is a minimum, and for what value $f$ is a maximum. We shall designate these two values (or sets of values) by the symbols $B_{\min}$ and $B_{\max}$, respectively.

Consider first the determination of $B_{\min}$. This is obtained from the conditions

$$\frac{\partial\bar{D}}{\partial B}=0;\quad \frac{\partial^2\bar{D}}{\partial B^2}>0.$$

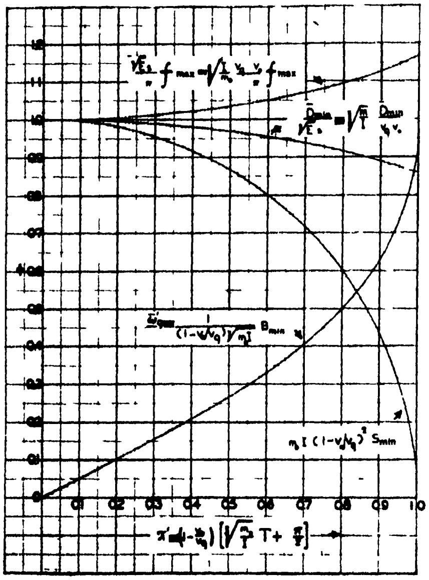


FIG. 6. Optimum performance curves for minimum deviation and maximum fidelity—proportional robot in constant tracking.

In general this value of $B_{\min}$ will depend on the time as well as the parameters, and to every time there will be associated a minimizing value $B_{\min}$. The value of $\bar{D}$ associated with $B_{\min}$ may be obtained by substituting into Eq. (1.06). $\bar{D}_{\min}$ and $B_{\min}$ as ordinates may both be plotted against the associated time $T$. Examples of such curves for proportional and on-off control robots obtained from the mathematical analysis in Part II and a subsequent paper are given in Figs. 6, 7, and 8. The curves are approximate only, and valid under the conditions for short period oscillations, i.e., for low damping $B$ relative to motor strength $m_0$. It is further shown in the mathematical analysis that real minimum values of $\bar{D}_{\min}$, and maximum values of $f_{\max}$, may be obtained only for certain ranges of $B$ and $T$. These ranges are diagrammed in Fig. 9. For generality the ranges of negative $B$ and $T$ are also given.

The actual value of $T$, and therefore $B_{\min}$, to be chosen depends of course on the requirements of the situation. In many cases ($\bar{D}_x=0$) it will be found that $\bar{D}_{\min}$ may be made as small as desired by choosing a sufficiently large $T$. But generally we do not wish this time to be unduly long. The most practical procedure to be followed for systems with adjustable $B$ is then probably this:

*For a given power source $m_0$, load inertia $I$, constant velocity of tracking $v_g$, and initial load velocity $v_0$, if we designate a deviation amplitude which is not to be exceeded, we determine a point on the ($\bar{D}_{\min}-T$) curve. By making the friction coefficient in the system have the value $B_{\min}$ associated with the abscissa of this point, we ensure that the desired diminution in $\bar{D}$ will occur in the shortest possible time. This time is given by the common abscissa.*†

---

† The formal proof that the conditions for

*$\bar{D}$ minimum with respect to $B$ at constant $T$*

are entirely equivalent to those for

*$T$ minimum with respect to $B$ at constant $\bar{D}$*

is easy. It follows immediately from the expression

$$\bar{D}-\bar{D}(B,T)=0$$

that

$$\left(\frac{\partial T}{\partial B}\right)_{\bar{D}}=-\frac{(\partial\bar{D}/\partial B)_T}{(\partial\bar{D}/\partial T)_B},$$

and from this the desired conclusion may be drawn except in the case where the denominator on the right-hand side also vanishes (i.e., in the steady state).

As indicated above, just as one may seek to find an optimum value of $B$ for minimizing the deviation, one may instead determine a $B_{\max}$ to maximize the fidelity fraction. $B_{\max}$ is determined to satisfy the conditions

$$\partial f/\partial B=0; \quad \partial^2 f/\partial B^2<0.$$

Plots of $f_{\max}$ and $B_{\max}$ curves analogous to the $\bar{D}_{\min}$ and $B_{\min}$ curves may be made and the following prescription carried out:

*For a given power source $m_0$, load inertia $I$, constant velocity of tracking $v_q$, and initial velocity $v_0$, if we designate a fidelity fraction which is to be exceeded, we determine a point on the $f_{\max}$ curve. By making the friction coefficient in the system have the value $B_{\max}$ associated with the abscissa of this point, we ensure that the desired increase in $f$ will occur in the shortest possible time. This time is given by the abscissa.*

Such curves as $\bar{D}_{\min}$–$B_{\min}$ and $f_{\max}$–$B_{\max}$ may conveniently be referred to as *curves for optimum performance.*

We may define the *sensitivity* of a maximum or minimum as *the relative change in the ordinate for a unit change in the abscissa* away from the extremum. It is proposed that this is a fairly relevant criterion in most cases where the local importance of a maximum or minimum is to be estimated. With application to the functions $\bar{D}(B)$ and $f(B)$ this yields

$$S_{\min}=\frac{1}{\bar{D}_{\min}}\frac{\partial^2\bar{D}}{\partial B^2}\bigg|_{B_{\min}} \qquad S_{\max}=-\frac{1}{f_{\max}}\frac{\partial^2 f}{\partial B^2}\bigg|_{B_{\max}},$$

where in each case the second derivative is evaluated at the extremum, i.e., where the first derivative vanishes. Once the extremum problems have been solved, plots of $S_{\min}$ and $S_{\max}$ similar to those described for $\bar{D}_{\min}$ and $f_{\max}$ may of course be made.

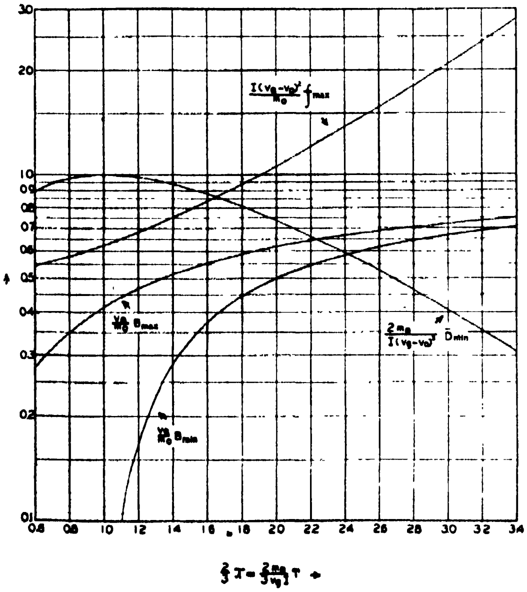


FIG. 7. Optimum performance curves for minimum deviation and maximum fidelity —on-off robot in constant tracking.

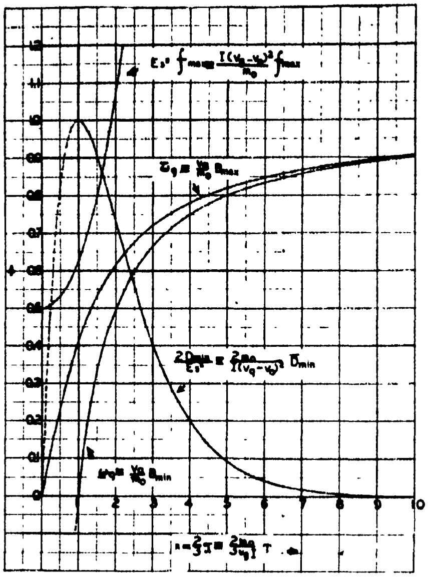


FIG. 8. Optimum performance curves on a semi-logarithmic plot—on-off robot in constant tracking.

So far, in the interest of clarity we have described the curves for optimum performance in general without specifying the restrictions which may be placed, by dimensional considerations, on the manner of occurrence of the various parameters. But it may be of *considerable practical value* to determine in general *upon what involvements of the parameters the optimum performance curves depend.* Various minimal sets of such variables may be found, with differing advantages.

Here we shall make a particular choice which brings out clearly the dependence on $B$: If the abbreviations

$$\frac{m_0 T}{v_g I}\equiv \mathcal{T}\qquad \frac{I v_g^2}{m_0}\equiv E\qquad 1-\frac{v_0}{v_g}\equiv s \qquad (\text{I}.07)$$

are introduced, it is readily seen that (I.05) and (I.06) may be rewritten

$$(D')\quad \bar{D}=\frac{E}{\omega_g^2}\Delta\left(\omega_g\mathcal{T},\ \frac{E}{\omega_g^2},\ \omega_g s,\ \omega_g\right) \qquad (\text{I}.08)$$

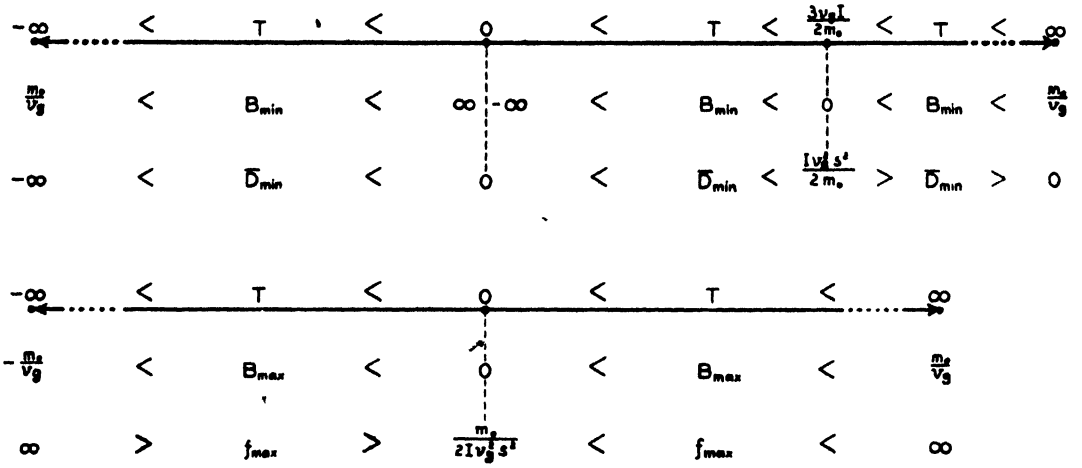
$$(f')\quad f=\frac{\omega_g^2}{E}f_\Delta\left(\omega_g\mathcal{T},\ \frac{E}{\omega_g^2},\ \omega_g s,\ \omega_g\right). \qquad (\text{I}.09)$$

It is clear from the definitions (I.07) of $\mathcal{T}$, $E$, and $s$ that they in no way involve $B$. The functions in $(D')$ and $(f')$ depend on $B$ only through

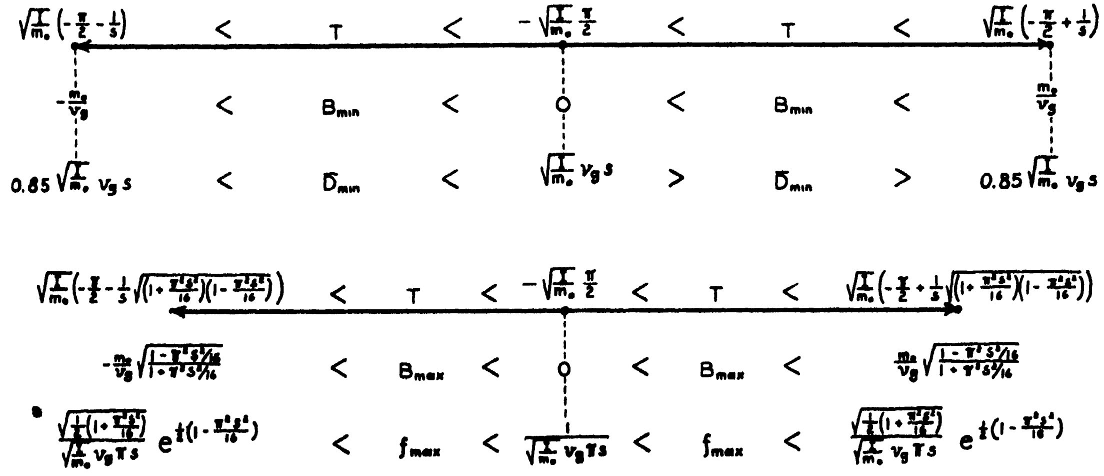
$$\omega_g\equiv\frac{B}{m_0}v_g.$$

Therefore all extremalizing values of $B$ may be obtained by setting

$$\frac{\partial\bar{D}}{\partial\omega_g}=0,\qquad \frac{\partial f}{\partial\omega_g}=0.$$


$-\infty < T < 0 < T < \frac{3v_gI}{2m_0} < T < \infty$
$\frac{m_0}{v_g} < B_{min} < \infty \,|\, -\infty < B_{min} < 0 < B_{min} < \frac{m_0}{v_g}$
$-\infty < \bar{D}_{min} < 0 < \bar{D}_{min} < \frac{Iv_g s^2}{2m_0} > \bar{D}_{min} > 0$
$-\infty < T < 0 < T < \infty$
$-\frac{m_0}{v_g} < B_{max} < 0 < B_{max} < \frac{m_0}{v_g}$
$\infty > f_{max} > \frac{m_0}{2Iv_g s^2} < f_{max} < \infty$

ON - OFF ROBOTS


$\sqrt{\frac{I}{m_0}}\left(-\frac{\pi}{2}-\frac{1}{s}\right) < T < -\sqrt{\frac{I}{m_0}}\frac{\pi}{2} < T < \sqrt{\frac{I}{m_0}}\left(-\frac{\pi}{2}+\frac{1}{s}\right)$
$-\frac{m_0}{v_g} < B_{min} < 0 < B_{min} < \frac{m_0}{v_g}$
$0.85\sqrt{\frac{I}{m_0}}v_g s < \bar{D}_{min} < \sqrt{\frac{I}{m_0}}v_g s > \bar{D}_{min} > 0.85\sqrt{\frac{I}{m_0}}v_g s$
$\sqrt{\frac{I}{m_0}}\left(-\frac{\pi}{2}-\frac{1}{s}\sqrt{\left(1+\frac{\pi^2s^2}{16}\right)\left(1-\frac{\pi^2s^2}{16}\right)}\right) < T < -\sqrt{\frac{I}{m_0}}\frac{\pi}{2} < T < \sqrt{\frac{I}{m_0}}\left(-\frac{\pi}{2}+\frac{1}{s}\sqrt{\left(1+\frac{\pi^2s^2}{16}\right)\left(1-\frac{\pi^2s^2}{16}\right)}\right)$
$-\frac{m_0}{v_g}\sqrt{\frac{1-\pi^2s^2/16}{1+\pi^2s^2/16}} < B_{max} < 0 < B_{max} < \frac{m_0}{v_g}\sqrt{\frac{1-\pi^2s^2/16}{1+\pi^2s^2/16}}$
$\frac{\sqrt{\frac{1}{2}\left(1+\frac{\pi^2s^2}{16}\right)}}{\sqrt{\frac{I}{m_0}}v_g\pi s}e^{\frac{1}{2}\left(1-\frac{\pi^2s^2}{16}\right)} < f_{max} < \frac{1}{\sqrt{\frac{I}{m_0}}v_g\pi s} < f_{max} < \frac{\sqrt{\frac{1}{2}\left(1+\frac{\pi^2s^2}{16}\right)}}{\sqrt{\frac{I}{m_0}}v_g\pi s}e^{\frac{1}{2}\left(1-\frac{\pi^2s^2}{16}\right)}$

PROPORTIONAL ROBOTS

FIG. 9. Domain of real $\bar{D}_{min}$ and $f_{max}$ with associated ranges of the time and the friction coefficient.

These equations yield

$$(\omega_g)_{\min} \quad (\omega_g)_{\max}$$
$$\bar{D}_{\min} \quad f_{\max}$$

and the sensitivities, etc., solely in terms of the quantities

$$\mathcal{T} \quad E \quad s$$

and no others. *Thus for any simple robot the optimum performance curves depend upon the time and physical parameters only in the three combinations defined in (I.07).* When quantities are expressed in terms of $\mathcal{T}$, $E$, and $s$, we shall say that they are "in $\mathcal{T}Es$ measure."

We shall now illustrate the use of the optimum performance curves by a practical example. Consider a system with the following parameters:

| | |
|---|---|
| motor strength | $m_0 = 1$ kw. $= 10^{10}$ erg/sec. |
| inertia | $I = 1000$ kg $= 10^6$ g |
| guide velocity | $v_g = 3.6$ km/hr. $= 100$ cm/sec. |
| initial load velocity | $v_0 = 1.8$ km/hr. $= 50$ cm/sec. |
| $E = 1$ | $s = \frac{1}{2}$. |

Suppose we wish to determine the optimum conditions for the deviation amplitude not to exceed 0.45 cm, i.e.,

$$\frac{\bar{D}_{\min}}{(E)^{\frac{1}{2}}s} = 0.9.$$

We find from Fig. 6 for proportional robots:

$$\mathcal{T}' = 0.86 \qquad \mathrm{T} = \left(\frac{I}{m_0}\right)^{\frac{1}{2}}\left(-\frac{\pi}{2} + \frac{\mathcal{T}'}{s}\right) = 0.0015 \text{ sec.}$$

$$\underline{\bar{\omega}}_g' = 0.57 \qquad B = 28 \times 10^6 \text{ dynes per cm/sec.}$$
$$= 28.6 \text{ kg per cm/sec.}$$

$$(E)^{\frac{1}{2}}\pi s f_{\max} = 1.11 \qquad f_{\max} = 0.707 \text{ cm}^{-1}.$$

Thus the optimal friction coefficient is such as to produce a drag of about 29 kg for a velocity 1 cm/sec. After 15 ten thousandths of a second the deviation amplitude will have fallen below 0.45 cm. To use the fidelity fraction, we consider the sample question: What fraction of a cycle does the system spend within a deviation equal to $\frac{1}{5}$ the deviation amplitude, i.e., within 0.09 cm of the correct position? The answer is $0.707 \times 2 \times 0.09 = 0.127 \simeq \frac{1}{8}$ of a cycle.

When we now consider an on-off robot with the same parameters we find we can no longer choose the minimizing deviation at 0.45 cm, since this already exceeds the upper limit for which $\bar{D}_{\min}$ is real. In fact we have a real minimum only when

$$\bar{D}_{\min} \leqslant \frac{I(v_g - v_0)^2}{2m_0} = 0.125 \text{ cm}$$

in this case. Taking $\bar{D}_{\min}$ equal to $\frac{1}{4}$ its value in the preceding example, i.e., at 0.1125 cm, we find

$$\frac{2\bar{D}_{\min}}{Es^2} = 0.9,$$

for which Fig. 7 yields:

$$\tfrac{2}{3}\mathcal{T} = 1.53 \qquad T = 0.023 \text{ sec.}$$

$$\underline{\omega}_g = 0.35 \qquad B_{\min} = 35.7 \text{ kg per cm/sec.}$$

$$\bar{\omega}_g = 0.54 \qquad B_{\max} = 55 \text{ kg per cm/sec.}$$

$$Es^2 f_{\max} = 0.81 \qquad f_{\max} = 3.24 \text{ cm}^{-1}.$$

The fidelity fraction in this case yields the result that the system will probably spend a fraction

$$3.24 \times 2 \times \frac{0.1125}{5} = 0.146 \text{ of a cycle}$$

within a deviation equal to $\frac{1}{5}$ the amplitude about the correct position.

## PART II. THEORY OF SIMPLE ROBOTS WITH PROPORTIONAL CONTROL

### A. Complete Solutions as Functions of Time Optimum Performance Curves

We shall now consider a particular type of control for which detailed analytical and graphical solutions will be given. This is the case of the simple robot with *proportional control* in which the restoring force is linearly proportional to the deviation, i.e.,

$$m = m_0(\theta_g - \theta).$$

The basic equation for the robot with usual load may then be written

$$I\frac{d^2\theta}{dT'^2} + B\frac{d\theta}{dT'} = m_0(\theta_g - \theta)$$

$$\left.\begin{aligned} \theta &= \theta_b \\ \frac{d\theta}{dT'} &= \omega_b \end{aligned}\right\} \text{at } T' = 0. \tag{II.01}$$

Introducing the dimensionless measures given by (I.02′) and taking the case of constant tracking, we have

$$\frac{d^2\Delta'}{dT'^2}+\frac{d\Delta'}{dT'}+\kappa\Delta'=\omega_g$$

$$\left.\begin{aligned}\Delta'=\Delta_b'=-\phi_b\\ \frac{d\Delta'}{dT'}=\omega_g-\omega_b\equiv\Omega_b\end{aligned}\right\}\text{at } T'=0. \tag{II.02}$$

Upon integration of (II.02) we get:

$$\Delta'=\frac{\omega_g}{\kappa}+(\phi_b+\omega_g/\kappa)e^{-T'/2}\frac{\sin\nu_0(T'-T_0')}{\sin\nu_0T_0'} \tag{II.02′}$$

with

$$T_0'=\frac{1}{\nu_0}\left[\pi-\sin^{-1}\frac{1}{[1+(-z_b/\nu_0)^2]^{\frac{1}{2}}}\right]$$

$$=\frac{1}{\nu_0}\left[\pi-\tan^{-1}\left(\frac{\nu_0}{-z_b}\right)\right]$$

$$=\frac{1}{\nu_0}\left[\frac{\pi}{2}-\tan^{-1}\frac{z_b}{\nu_0}\right]$$

where we have introduced the convenient abbreviations

$$z_b\equiv\frac{\Omega_b}{\phi_b+\omega_g/\kappa}-\tfrac{1}{2}\equiv\frac{I}{B}\frac{v_g-v_b}{\theta_b+Bv_g/m_0}-\tfrac{1}{2}$$
$$\nu_0\equiv(\kappa-\tfrac{1}{4})^{\frac{1}{2}}\equiv\frac{(4m_0I-B^2)^{\frac{1}{2}}}{2B} \tag{II.03}$$

It is clear that the mean steady-state deviation in this type of system with proportional control is then

$$\Delta_\infty'=\frac{\omega_g}{\kappa}\quad D_\infty'=\frac{B}{m_0}v_g$$

and is just *the deflection which evokes the restoring force required to drag the system along against friction at the velocity $v_g$*. This is of course only a special case of the general result for $\Delta_\infty'$ described in the preceding section. For the time dependent deviation from $\Delta_\infty'$ there are three cases:

*Case I*: $\nu_0^2>0$. *Oscillatory Approach to Steady State*

$$\Delta=(\phi_b+\omega_g/\kappa)e^{-T'/2}\frac{\sin\nu_0(T'-T_0')}{\sin\nu_0T_0'}.$$

As $v_0$ approaches 0, $T_0'$ may be represented approximately by

$$T_0'\simeq\frac{\pi}{\nu_0}+\frac{1}{z_b}\quad\text{if}\quad z_b<0,$$

$$T_0'\simeq\frac{1}{z_b}\qquad\text{if}\quad z_b>0.$$

The deviation amplitude in the steady state $\Delta_\infty$ is zero, and the number of transient cycles $n_\infty$ and time $T_\infty$ required to reach the steady state are both infinite.

*Case II*: $\nu_0^2=0$. *Critically Damped Approach to Steady State*

$$\Delta=(\phi_b+\omega_g/\kappa)e^{-T'/2}\left(\frac{T'}{T_0'}-1\right)$$

where

$$T_0'=\infty\quad\text{if}\quad z_b<0,$$

$$T_0'=\frac{1}{z_b}\quad\text{if}\quad z_b>0.$$

Thus we see that the number of cycles $n_\infty$ before the asymptotically approached steady state is either 0 or ½, depending on whether $z_b$ is less than or greater than zero. In the latter case Δ becomes zero once at $T_0'$, and overshoots the mark, later subsiding asymptotically to zero again. Examining the structure of $z_b$, we see that this occurs under the following sets of circumstances:

$I/B$ sufficiently large
$v_g-v_b$ sufficiently large
$\theta_b$ small, $m_0/Bv_g$ sufficiently large.

*Case III*: $\nu_0^2<0$. *Overdamped Approach to Steady State*

$$\Delta=(\phi_b+\omega_g/\kappa)e^{-T'/2}\frac{\sinh|\nu_0|(T'-T_0')}{\sinh|\nu_0|T_0'}$$

where

$$T_0' = \frac{1}{|\nu_0|}\sinh^{-1}\frac{1}{[(-z_b/\nu_0)^2-1]^{\frac{1}{2}}}$$

$$= -\frac{1}{|\nu_0|}\tanh^{-1}\frac{|\nu_0|}{-z_b}$$

$$= \ln\left|\frac{z_b+|\nu_0|}{z_b-|\nu_0|}\right|^{1/2|\nu_0|}.$$

It appears that $T_0'$ has a real positive finite value only for $z_b > |\nu_0|$. Thus the smallest possible real values of $T_0'$ and the number $n_\infty$ of transient cycles during the approach to the steady state of zero deviation are governed by the rule

$$T_0' \text{ finite} \quad n_\infty = \tfrac{1}{2} \quad \text{if } z_b > |\nu_0|$$
$$T_0' = \infty \quad n_\infty = 0 \quad \text{if } z_b < |\nu_0|.$$

This is to be compared with the corresponding condition in Case II: A schedule of the possible real $T_0'$ in the various cases, together with an analysis, by means of the unit circle in the trigonometric plane, of the limits approached, is given in Table IV.

For fixed $z_b > 0$, it is clear from the nature of the function $(1/x)\tanh^{-1}x$ that $T_0'$ has its minimum value at $\nu_0 = 0$, i.e., for critical damping. Thus, *if the objective is to attain* $T_0'$ *as quickly as possible, we shall certainly strive to keep the damping below the critical value.*

We pass now to the consideration of the amplitude time dependence in the various cases. The most interesting circumstances are those in which there is oscillation as in Case I, or when there is overshoot in Cases II and III.

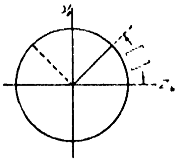
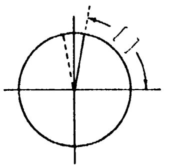
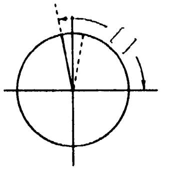
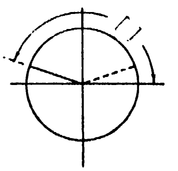
*Case* I: If we shift to the time scale

$$T = T' - T_0'$$

$$\Delta = (\phi_b + \omega_g/\kappa)e^{-(T+T_0')/2}\frac{\sin\nu_0 T}{\sin\nu_0 T_0'}$$

TABLE IV. Smallest real $T_0'$ as a function of $\nu_0$ and $z_b$ for the proportional robot.

$\longleftarrow \nu_0^2$

| | $\nu_0$ Real: oscillation<br>$T_0' = \frac{1}{\nu_0}\left\lvert \pi - \tan^{-1}\frac{\nu_0}{-z_b}\right\rvert = \frac{[\ ]}{\nu_0}$ | | $\nu_0 \to$ zero<br>Critical damping | zero $\leftarrow \lvert\nu_0\rvert$ | $\nu_0$ Imaginary: overdamping |
|---|---|---|---|---|---|
| $z_b$ plus | |  | $T_0' = \frac{1}{z_b}$ | $T_0' = \frac{1}{z_b}$ | if $z_b > \lvert\nu_0\rvert$:<br>$T_0' = \frac{1}{\lvert\nu_0\rvert}\tanh^{-1}\frac{\lvert\nu_0\rvert}{z_b}$<br>if $z_b \leq \lvert\nu_0\rvert$:<br>$\infty$ |
| ↓ | | | | | |
| $z_b$ zero | $T_0' = \frac{1}{\nu_0}\left[\pi - \frac{\pi}{2}\right]$<br>$= \pi/2\nu_0$ |  | $\frac{\pi}{2\nu_0} \to \infty$ | $\infty$ | $\infty$ |
| ↑<br>$z_b$ | | | | | |
| $z_b$ zero | $T_0' = \pi/2\nu_0$ |  | $\frac{\pi}{2\nu_0} \to \infty$ | $\infty$ | $\infty$ |
| ↑ | | | | | |
| $z_b$ minus | • |  | $\frac{\pi}{\nu_0} \to \infty$ | $\infty$ | $\infty$ |

$z_b$ can be plus or minus.
$\nu_0$ can be real or imaginary.

we may represent the beginning of each cycle by

$$\mathrm{T}_n=\frac{2n\pi}{\nu_0}\qquad (n=0,\,1,\,2,\,\cdots)$$

and the middle (node) by

$$\mathrm{T}_{n+\frac{1}{2}}=\frac{(2n+1)\pi}{\nu_0}=\mathrm{T}_n+\frac{\pi}{\nu_0}.$$

The times when the maximum and minimum deviations occur in the cycle are found by examining the zeros of the velocity

$$\Omega\equiv\frac{d\Delta}{d\mathrm{T}}=\frac{\phi_b+\omega_g/\kappa}{\sin\nu_0\mathrm{T}_0'}e^{-(\mathrm{T}+\mathrm{T}_0')/2}$$

$$\times\left[-\tfrac{1}{2}\sin\nu_0\mathrm{T}+\nu_0\cos\nu_0\mathrm{T}\right]$$

to be

$$\bar{\mathrm{T}}_n=\mathrm{T}_n+\frac{1}{\nu_0}\tan^{-1}2\nu_0$$

$$\underline{\mathrm{T}}_n=\mathrm{T}_n+\frac{1}{\nu_0}\tan^{-1}2\nu_0+\frac{\pi}{\nu_0}.$$

In particular

$$\bar{\mathrm{T}}_0=\frac{1}{\nu_0}\tan^{-1}2\nu_0.$$

The maximum amplitude of deviation in the $n$-th cycle, or the amplitude-time function, is then:

$$\bar{\Delta}_n=\bar{\Delta}_0e^{-\mathrm{T}_n/2},$$

where $\Delta_0$, the maximum deviation amplitude in the first cycle, is

$$\bar{\Delta}_0=(\phi_b+\omega_g/\kappa)e^{-(\bar{\mathrm{T}}_0+\mathrm{T}_0')/2}\,\frac{\sin\nu_0\bar{\mathrm{T}}_0}{\sin\nu_0\mathrm{T}_0'}.$$

But

$$\bar{\mathrm{T}}_0+\mathrm{T}_0'=\frac{1}{\nu_0}\left[\pi-\tan^{-1}\frac{\nu_0}{-z_b}+\tan^{-1}2\nu_0\right]$$

$$=\frac{1}{\nu_0}\left[\pi-\tan^{-1}\left(\nu_0\frac{2z_b+1}{-z_b+2\nu_0^2}\right)\right],$$

and

$$\frac{\sin\kappa_0\bar{\mathrm{T}}_0}{\sin\nu_0\mathrm{T}_0'}=\left(\frac{1+(z_b/\nu_0)^2}{1+(1/2\nu_0)^2}\right)^{\frac{1}{2}},$$

so that

$$\bar{\Delta}_0=(\phi_b+\omega_g/\kappa)\left(\frac{\nu_0^2+z_b^2}{\nu_0^2+\frac{1}{4}}\right)^{\frac{1}{2}}$$

$$\times\exp-\frac{1}{2\nu_0}\left[\pi-\tan^{-1}\left(\nu_0\frac{2z_b+1}{-z_b+2\nu_0^2}\right)\right].$$

*Case* II, $z_b>0$: The time when the maximum deviation amplitude occurs is here given by

$$\mathrm{T}=2\quad\text{or}\quad \mathrm{T}'=\mathrm{T}_0'+2=\frac{1}{z_b}+2,$$

and the associated maximum deviation amplitude is

$$\bar{\Delta}=2z_b(\phi_b+\omega_g/\kappa)e^{-(1+1/2z_b)}.$$

*Case* III, $z_b>|\nu_0|$: With respect to the shifted time scale, $\Delta$ may here be written:

$$\Delta=(\phi_b+\omega_g/\kappa)((z_b/\nu_0)^2-1)^{\frac{1}{2}}\left[\frac{z_b-|\nu_0|}{z_b+|\nu_0|}\right]^{1/4|\nu_0|}$$

$$e^{-\mathrm{T}/2}\sinh|\nu_0|\mathrm{T}.$$

The time when the maximum amplitude of overshoot is reached is found by examining the zero of $d\Delta/d\mathrm{T}'$, to be

$$\bar{\mathrm{T}}=\frac{1}{\nu_0}\tanh^{-1}2|\nu_0|=\ln\left[\frac{1+2|\nu_0|}{1-2|\nu_0|}\right]^{1/2|\nu_0|}$$

which, upon insertion into the expression for $\Delta$, yields the maximum amplitude

$$\bar{\Delta}=(\phi_b+\omega_g/\kappa)\left(\frac{z_b^2-|\nu_0|^2}{\frac{1}{4}-|\nu_0|^2}\right)^{\frac{1}{2}}$$

$$\times\left[\frac{1-2|\nu_0|}{1+2|\nu_0|}\,\frac{z_b-|\nu_0|}{z_b+|\nu_0|}\right]^{1/4|\nu_0|}.$$

It is clear from the above expressions that in order that there shall actually be finite real values for T and $\Delta$, it is also necessary that

$$|\nu_0|<\tfrac{1}{2}\ (\nu_0\ \text{pure imaginary}).$$

But this is merely the condition that

$$\kappa\equiv\frac{m_0I}{B^2}>0$$

i.e., *in order that the time and amplitude of overshoot may not increase indefinitely it is necessary that the restoring force factor be positive.* This requirement is otherwise obvious on physical grounds.

We shall now determine the fidelity fraction for Case I, which is the only case in which the concept has much interest. The expression for $\Omega$ given above may be rewritten

$$\Omega(\mathrm{T}) = -\Delta_b\nu_0(1+(z_b/\nu_0)^2)^{\frac{1}{2}}e^{-(\mathrm{T}+\mathrm{T}_0')/2}\cos\nu_0\mathrm{T} - \tfrac{1}{2}\Delta(\mathrm{T}),$$

and if we evaluate this at $\mathrm{T}=\mathrm{T}_n$ and $\mathrm{T}=\mathrm{T}_{n+\frac{1}{2}}$, we find

$$\Omega(\mathrm{T}_n) = \nu_0\bar{\Delta}_n e^{\bar{\mathrm{T}}_0/2}(1+1/4\nu_0^2)^{\frac{1}{2}}$$

$$= \nu_0 \exp\frac{1}{2\nu_0}\tan^{-1}2\nu_0(1+1/4\nu_0^2)^{\frac{1}{2}}\bar{\Delta}_n$$

$$\Omega(\mathrm{T}_{n+\frac{1}{2}}) = -\Omega(\mathrm{T}_n)e^{-\pi/2\nu_0}.$$

Inserting these expressions into the formulas for $f_\Delta$, and remembering that the period $\pi_n$ is given by the constant $2\pi/\nu_0$, we have

$$f_\Delta = \frac{1}{2\pi}\exp-\frac{1}{2\nu_0}\tan^{-1}2\nu_0(1+e^{\pi/2\nu_0})\frac{1}{\bar{\Delta}_n(1+1/4\nu_0^2)^{\frac{1}{2}}}.$$

*Thus, except for a constant factor, the fidelity fraction varies inversely with the deviation amplitude.*

In ordinary dimensional and $\mathcal{T}Es$ measure (see definitions (I.11) Section (C) Part I) the amplitude-time function may be written.

$$\bar{D} = \bar{D}_0 e^{-(B/2I)T} \equiv \bar{D}_0 e^{(\omega_g/2)\mathcal{T}}$$

and the fidelity fraction is:

$$f = f_0 e^{(B/2I)T} \equiv f_0 e^{(\omega_g/2)\mathcal{T}},$$

where

$$\bar{D}_0 = \left(\theta_b + \frac{B}{m_0}v_g\right)\left(\frac{\nu_0^2+z_b^2}{\nu_0^2+\frac{1}{4}}\right)^{\frac{1}{2}} \times \exp-\frac{1}{2\nu_0}\left[\pi - \tan^{-1}\left(\nu_0\frac{2z_b+1}{-z_b+2\nu_0^2}\right)\right],$$

$$f_0 = \frac{1}{2\pi}\exp-\frac{1}{2\nu_0}\tan^{-1} \times 2\nu_0(1+e^{\pi/2\nu_0})\frac{1}{\bar{D}_0(1+1/4\nu_0^2)^{\frac{1}{2}}}.$$

For simplicity we shall set $\theta_b=0$ in the following analysis. The definitions (I.03) may then be written in $\mathcal{T}Es$ measure

$$\nu_0 = \left(\frac{E}{\omega_g^2} - 1\right)^{\frac{1}{2}}$$
$$z_b = \frac{Es}{\ } - \tfrac{1}{2} \equiv s(\nu_0^2+1) - \tfrac{1}{2}. \tag{II.04}$$

The argument of the arc-tangent in the exponent of $\bar{D}_0$ then takes the form

$$\nu_0\frac{2z_b+1}{2\nu_0^2-z_b} = \nu_0\frac{2s(\nu_0^2+1)}{\nu_0^2(2-s)-s+\frac{1}{2}}$$

We shall examine in detail only the approximation $\nu_0>>1$ corresponding to short period oscillations, or low damping $B$ relative to the square root of the product of motor strength $m_0$ by inertia $I$. For this limit the arc-tangent approaches $\pi/2$ and we may write

$$\bar{D}_0 \simeq \omega_g(1+s^2\nu_0^2)^{\frac{1}{2}}e^{-(\pi/4\nu_0)}$$

or, making use of the fact that in this approximation

$$\nu_0 = \frac{(E)^{\frac{1}{2}}}{\omega_g},$$

we may write

$$\bar{D} = (\omega_g^2+Es^2)^{\frac{1}{2}}\exp-\left(\frac{\pi}{4(E)^{\frac{1}{2}}}+\frac{\mathcal{T}}{2}\right)\omega_g$$

$$= \left(\left(\frac{Bv_g}{m_0}\right)^2 + \frac{Iv_g^2}{m_0}\left(1-\frac{v_0}{v_g}\right)^2\right)^{\frac{1}{2}} \times \exp\left(-\frac{B}{2I}T - \frac{\pi B}{4(m_0I)^{\frac{1}{2}}}\right).$$

We pause here for a moment to draw some general conclusions. It is clear from the expression that *the deviation amplitude at any given time will be smaller the smaller the guide velocity $v_g$ and its difference from the initial velocity of the system $v_0$, the smaller the inertia of the load $I$, and the greater the strength of the power source $m_0$* (above the maximizing value $m_0^*=8(I)^{\frac{1}{2}}/\pi B$). As mentioned before, the inertia of the load, the strength of the power source, etc., will often be fixed by other considerations. The most controllable factor is likely to be the friction coefficient $B$ as contained in the parameter $\omega_g$. We now proceed to determine for what values of $\omega_g$, $\bar{D}$ is a minimum, the value of this minimum, and its sensitivity (defined in the introduction).

It will be convenient to introduce the abbreviations

$$\mathcal{T}' = (E)^{\frac{1}{2}} s\left(\tau + \frac{\pi}{2(E)^{\frac{1}{2}}}\right)$$

$$\equiv \left(1 - \frac{v_0}{v_g}\right)\left(\left(\frac{m_0}{I}\right)^{\frac{1}{2}} T + \frac{\pi}{2}\right)$$

$$\omega_g' \equiv \frac{\omega_g}{(E)^{\frac{1}{2}} s} \equiv \frac{B}{(m_0 I)^{\frac{1}{2}}} \frac{v_g}{v_g - v_0}$$

as measures of the time and guide velocity in this case. With these notations we may write

$$\bar{D} = (E)^{\frac{1}{2}} s (1 + \omega_g'^2)^{\frac{1}{2}} e^{-(\omega_g'/2)\mathcal{T}'}.$$

From the formula for $\bar{D}$ we find, consecutively,

$$\frac{\partial \bar{D}}{\partial \omega_g'} = \bar{D}\left[\frac{\omega_g'}{1 + \omega_g'^2} - \frac{\mathcal{T}'}{2}\right],$$

$$\frac{\partial^2 \bar{D}}{\partial \omega_g'^2} = \frac{1}{\bar{D}}\left(\frac{\partial \bar{D}}{\partial \omega_g'}\right)^2 + \bar{D}\frac{1 - \omega_g'^2}{1 + \omega_g'^2}.$$

Thus

$$(\omega_g')_{\min} \equiv \underline{\omega}_g' = \frac{1 - (1 - \mathcal{T}'^2)^{\frac{1}{2}}}{\mathcal{T}'} \quad \omega_g'^2 = \frac{2\omega_g'}{\mathcal{T}'} - 1$$

$$\bar{D}_{\min} = (E)^{\frac{1}{2}} s \left(\frac{2\underline{\omega}_g'}{\mathcal{T}'}\right)^{\frac{1}{2}} e^{-(\omega_g'/2)\mathcal{T}'}$$

$$S_{\min} \equiv \frac{1}{\bar{D}_{\min}} \frac{\partial^2 \bar{D}}{\partial \omega_g'^2} \left(\frac{d\omega_g'}{dB}\right)^2$$

$$= \frac{1}{m_0 I (1 - v_0/v_g)^2} \left(\frac{\mathcal{T}'}{\underline{\omega}_g'} - 1\right) > 0,$$

where we have indicated the necessary condition that the extreme value represent a minimum by the inequality on $S$.

In the neighborhood of $\mathcal{T}' = 0$ we have

$$(\omega_g')_{\min} \simeq \frac{\mathcal{T}'}{2}$$

$$\frac{\bar{D}_{\min}}{(E)^{\frac{1}{2}} s} \simeq e^{-(\mathcal{T}'/2)^2} \tag{II.05}$$

$$m_0 I (1 - v_0/v_g)^2 S_{\min} \simeq 1,$$

and in the neighborhood of $\mathcal{T}' = 1$

$$\underline{\omega}_g' \to 1 \quad \frac{d\underline{\omega}_g'}{d\mathcal{T}'} \to \infty \quad \text{(cusp)}$$

$$\frac{\bar{D}_{\min}}{(E)^{\frac{1}{2}} s} \to (2)^{\frac{1}{2}} e^{-\frac{1}{2}} \sim 0.85 \tag{II.06}$$

$$S_{\min} \to 0.$$

This limiting behavior, as well as the variation throughout the interval, is exhibited in Fig. 2, where $\omega_g'$, $\bar{D}_{\min}/(E)^{\frac{1}{2}} s$, and $m_0 I (1 - v_0/v_g)^2 S_{\min}$ are all plotted as functions of the time variable $\mathcal{T}'$. Beyond $\mathcal{T}' = 1$ there is no real minimum for $\bar{D}$.

In similar fashion to the above we may carry out the analysis for the fidelity fraction. We find

$$f_0 = \frac{1}{\pi} \frac{\cosh \frac{\pi}{4(E)^{\frac{1}{2}}} \omega_g}{\bar{D}_0} \equiv \frac{1}{\pi} \frac{\cosh \frac{\pi}{4} s \omega_g'}{\bar{D}_0},$$

so that

$$f = \frac{1}{\pi} \frac{\cosh \frac{\pi}{4} s \omega_g'}{\bar{D}}.$$

Making use of the preceding analysis of $\bar{D}$, we find

$$\frac{\partial f}{\partial \omega_g'} = f\left[\frac{\pi s}{4} \tanh \frac{\pi s}{4} \omega_g' - \frac{\omega_g'}{1 + \omega_g'^2} + \frac{\mathcal{T}'}{2}\right]$$

$$\frac{\partial^2 f}{\partial \omega_g'^2} = \frac{1}{f}\left(\frac{\partial f}{\partial \omega_g'}\right)^2 + f\left[\left(\frac{\pi s}{4}\right)^2 \operatorname{sech}^2 \frac{\pi s}{4} \omega_g' - \frac{1 - \omega_g'^2}{1 + \omega_g'^2}\right]. \tag{II.07}$$

If we set

$$\frac{\pi s}{4} \omega_g' \simeq 0$$

corresponding to the inequality $E >> \omega_g^2$, we may rewrite these expressions in the simplified form

$$\frac{\partial f}{\partial \omega_g'} = -f\left[\frac{\omega_g'}{1 + \omega_g'^2} - \frac{\mathcal{T}'}{2}\right]$$

$$\frac{\partial^2 f}{\partial \omega_g'^2} = \frac{1}{f}\left(\frac{\partial f}{\partial \omega_g'}\right)^2 + f\left[\left(\frac{\pi s}{4}\right)^2 - \frac{1 - \omega_g'^2}{1 + \omega_g'^2}\right].$$

From these we may deduce that to this approximation $(\underline{\omega}_g')_{\max}$ *is equal to* $(\omega_g')_{\min}$. Thus

$$(\omega_g')_{\max}=\underline{\omega}_g'=\frac{1-(1-\mathcal{T}'^2)^{\frac{1}{2}}}{\mathcal{T}'}$$

$$f_{\max}=\frac{1}{\pi\bar{D}_{\min}}=\frac{1}{\pi(E)^{\frac{1}{2}}s}\left(\frac{\mathcal{T}'}{2\underline{\omega}_g'}\right)^{\frac{1}{2}}e^{(\omega_g'/2)\mathcal{T}'}$$

$$m_0I(1-v_0/v_g)^2S_{\max}=\frac{\pi^2s^2}{16}+1-\frac{\mathcal{T}'}{\omega_g'}\leq 0.$$

$\pi(E)^{\frac{1}{2}}s\,f_{\max}$ is also plotted in Fig. 6. We note that the only difference in the conditions determining $f_{\max}$ and $\bar{D}_{\min}$ here is that the maximizing condition holds in a more restricted range, i.e.,

$$\frac{\underline{\omega}_g'}{\mathcal{T}'}\leq\frac{1}{1+(\pi s/4)^2},$$

which, together with the preceding equations, determines the following inequalities for the domain in which $f$ is still maximized

$$-\left(\frac{1-\pi^2s^2/16}{1+\pi^2s^2/16}\right)^{\frac{1}{2}}\leq\omega_g'\leq\left(\frac{1-\pi^2s^2/16}{1+\pi^2s^2/16}\right)^{\frac{1}{2}}$$

$$-[(1+\pi^2s^2/16)(1-\pi^2s^2/16)]^{\frac{1}{2}}\leq\mathcal{T}'$$
$$\leq[(1+\pi^2s^2/16)(1-\pi^2s^2/16)]^{\frac{1}{2}}$$

$$\frac{1}{(E)^{\frac{1}{2}}\pi s}\leq f_{\max}\leq\frac{1}{(E)^{\frac{1}{2}}\pi s}[\tfrac{1}{2}(1+\pi^2s^2/16)]^{\frac{1}{2}}e^{\frac{1}{2}(1-\pi^2s^2/16)}.$$

The formulas given here and above are the base for the general summary diagram (Fig. 9) at the end of the Part I indicating the domains of the variables in which we get real extrema for $\bar{D}_{\min}$ and $f_{\max}$.

### B. PATH CURVES IN THE POSITION-VELOCITY PLANE

For future use[1,2] we shall now work out the result of eliminating the time between the expression (II.02′) and its time derivative $\Omega\equiv d\Delta'/d\tau$ which satisfies the relation

$$\Omega+\tfrac{1}{2}(\Delta'-\omega_g/\kappa)=-\nu_0(\Delta_b'-\omega_g/\kappa)$$
$$\times e^{-\mathrm{T}'/2}\frac{\cos\nu_0(\mathrm{T}'-\mathrm{T}_0')}{\sin\nu_0\mathrm{T}_0'}.\qquad\text{(II.02'')}$$

Making the substitutions (which correspond to an "affine" transformation in the $\Omega-\Delta$ plane)

$$\Delta\equiv\Delta'-\omega_g/\kappa\quad y\equiv\Omega+\tfrac{1}{2}(\Delta'-\omega_g/\kappa),$$

with some rearrangement we may write (II.02′) and (II.02″):

$$\nu_0\Delta=-((\nu_0\Delta_b)^2+y_b^2)^{\frac{1}{2}}e^{-\mathrm{T}'/2}\sin\nu_0(\mathrm{T}'-\mathrm{T}_0'),$$
$$y=-((\nu_0\Delta_b)^2+y_b^2)^{\frac{1}{2}}e^{-\mathrm{T}'/2}\cos\nu_0(\mathrm{T}'-\mathrm{T}_0'),$$

from which we may find

$$\mathrm{T}'=\mathrm{T}_0'+\frac{1}{\nu_0}\tan^{-1}\frac{\nu_0\Delta}{y}=\frac{1}{\nu_0}\left[\tan^{-1}\frac{z}{\nu_0}-\tan^{-1}\frac{z_b}{\nu_0}\right],$$

where we have introduced the "clock-hand variable"

$$z\equiv-y/\Delta\equiv-(\Omega/\Delta+\tfrac{1}{2}).$$

Thus

$$(\nu_0\Delta)^2+y^2=[(\nu_0\Delta_b)^2+y_b^2]$$
$$\times\exp-\frac{1}{\nu_0}\left[\tan^{-1}\frac{z}{\nu_0}-\tan^{-1}\frac{z_b}{\nu_0}\right]\qquad\text{(II.08)}$$

which becomes in polar coordinates the equation of a right-handed (clockwise) equiangular spiral

$$\rho=\rho_be^{-(\psi-\psi_b)/2\nu_0}$$

with

radius vector $\rho\equiv((\nu_0\Delta)^2+y^2)^{\frac{1}{2}}$
(clockwise) angle $\psi\equiv\tan^{-1}(z/\nu_0)$.

Thus, in the case of underdamping (Case I) *the "path" of the linear robot in the phase-plane ($\Omega-\Delta'$ plane) is an affine distortion of a right-handed logarithmic spiral with center at $\omega_g/\kappa$ (or $\Delta=0$).* For the further analysis it is desirable to elucidate the meaning of what we have called the "clock-hand variable" $z$. We note that, in the $\Omega-\Delta$ plane, the successive values of $z$ correspond to straight lines through the origin arranged in clockwise increasing order. $z$, which equals the

negative slope $-\frac{1}{2}$

of these lines, goes from $-\infty$ to $+\infty$ twice in making one complete circuit; once in going from the $+\Omega$ axis to the $-\Omega$ axis, and a second time in going from the $-\Omega$ axis back to the $+\Omega$ axis. The zero of $z$ is not the $\Delta$ axis but rather the line

$$\Omega/\Delta=-\tfrac{1}{2}.$$

In the case of overdamping $\nu_0^2<0$ (Case III) we may write $\nu_0=i|\nu_0|$ and our basic Eq. (II.08) goes over to the form

$$y^2-|\nu_0|^2\Delta^2=(y_b^2-|\nu_0|^2\Delta_b^2)$$
$$\times\left|\frac{z_b-|\nu_0|}{z_b+|\nu_0|}\,\frac{z+|\nu_0|}{z-|\nu_0|}\right|^{1/2|\nu_0|}. \quad \text{(II.08')}$$

We see that all the curves beginning in the region $z_b<-|\nu_0|$ (i.e., the right-upper or left-lower portion of the plane) go in to zero touching along the line $z=-|\nu_0|$. All curves beginning in the region $z>|\nu_0|$ (i.e., right-lower or left-upper portion of the plane) overshoot $\Delta=0$, crossing the $\Delta$ axis at the points

$$\Omega=\pm[y_b^2-|\nu_0|^2\Delta_b^2]\left|\frac{z_b-|\nu_0|}{z_b+|\nu_0|}\right|^{1/4|\nu_0|}.$$

Finally all curves beginning in the intermediate region

$$-|\nu_0|<z_b<|\nu_0|$$

move counter-clockwise into the origin, touching the line $z=-|\nu_0|$ like all the others. If extrapolated backward ($z$ fixed, $z_b$ shifting) these lines are all asymptotically parallel to the line $z=|\nu_0|$.

In the case of critical damping $\nu_0=0$ (Case II), the equation of the path-curve takes the form

$$y^2\equiv(\Omega+\Delta\ 2)^2=(\Omega_b+\Delta_b/2)^2e^{1/z-1/z_b}, \quad \text{(II.08'')}$$

and the intermediate region of Case III has shrunk down to the line $y=0$ or $\Omega=-\Delta/2$.

It is interesting and important to note that if one takes the limit of (II.08') as the restoring force constant $\kappa\to0$ and $|\nu_0|\to\frac{1}{2}$, one obtains in the limit

$$\Omega=\Omega_b,$$

which is definitely *not* the general solution in this case.[2] This situation, where the limit of the solution when a coefficient *approaches zero* is not the same as the solution when the coefficient *is zero*, seems to be not uncommon in differential equations. Knowing that it occurs, one must then make special examination of the case at the limit. This case is contained in the analysis in the succeeding paper dealing with on-off systems.

# Computation of the Solutions of $(1+2\epsilon\cos 2z)y''+\theta y=0$; Frequency Modulation Functions

N. W. McLachlan
*London, England*

(Received January 9, 1947)

This equation pertains to a loss-free frequency modulation circuit. Floquet's theory is applied to obtain stable solutions when $\theta$, $\epsilon$ are outside the range of the approximation solution given by Carson in 1922. A method is described by means of which solutions may be computed to any degree of accuracy. The necessary formulae associated with the solution are derived, and convergence is discussed. A numerical example, illustrating the procedure to be followed, is worked out in detail. An appropriate normalization of the solutions is suggested in order to obtain standard frequency modulation functions. It is shown how the analysis may be extended to cover the case of the equation

$$d^2Q/dt^2+(R/L)(dQ/dt)+[Q/LC_0(1+2\epsilon\cos 2\omega_1 t)]=0,$$

which applies to a frequency modulation circuit having constant resistance $R$.

## 1. INTRODUCTION

THIS equation occurs in the theory of a method used for transmitting signal characters in radio broadcasting. The method is known as frequency modulation, and it is also the basis of the warble tone employed to reduce standing wave effects in acoustical test rooms. In the circuit shown schematically in Fig. 1A, the capacitance $C$ is a periodic function of the time $t$, being represented by the formula $C=C_0(1+2\epsilon\cos 2\omega_1 t)$, where $C_0$ and $\epsilon$ are parameters with $|\epsilon|<0.5$. If $Q$ is the quantity of electricity on the capacitance at any instant, the differential equation for the circuit is

$$(d^2Q/dt^2)+Q/LC_0(1+2\epsilon\cos 2\omega_1 t)=0. \qquad (1)$$

By writing $\omega_0{}^2=1/LC_0$, $\theta=(\omega_0/\omega_1)^2$, $\omega_1 t=z$, $y=Q$, (1) becomes

$$d^2y/dz^2+\theta y/(1+2\epsilon\cos 2z)=0. \qquad (2)$$

Expressed in this way, we see that if $|\epsilon|\geqslant 0.5$, $(1+2\epsilon\cos 2z)$ has zeros, and, therefore, $\theta/(1+2\epsilon\cos 2z)$ has corresponding infinities. These are avoided by imposing the restriction $|\epsilon|<0.5$. In radio broadcasting $|\epsilon|<<<0.5$, and to an adequate approximation (2) may be written

$$y''+(a\dot{-}2q\cos 2z)y=0, \qquad (3)$$

where $a=\theta$, and $q=\theta\epsilon$. Then (3) is the canonical form of Mathieu's differential equation. When (a), $a$ is very large and positive, (b) $a>>>q>0$, an approximate solution of (3) is given by

$$y=C\sum_{-\infty}^{\infty}J_r(q/2a^{\frac{1}{2}})\cos[(a^{\frac{1}{2}}-2r)z-\psi], \qquad (4)$$

$$=C\sum_{-\infty}^{\infty}J_r(\omega_0\epsilon/2\omega_1)\cos[(\omega_0-2r\omega_1)t-\psi], \qquad (5)$$

$C$ and $\psi$ being arbitrary constants. This form of solution was obtained by Carson in 1922.[1] Typical values of the various parameters in radio frequency modulation are: $f_0=\omega_0/2\pi=5\times 10^7$ c.p.s., $f_1=\omega_1/2\pi=5\times 10^3$ c.p.s., $f_0\epsilon=\omega_0\epsilon/2\pi$, $=\Delta f=\Delta\omega/2\pi=10^4$ c.p.s., this being the frequency deviation on either side of the central frequency $f_0$. Thus $\epsilon=2\times 10^{-4}$, so in (2) it is justifiable to take $(1+2\epsilon\cos 2z)^{-1}\simeq(1-2\epsilon\cos 2z)$. We also have $a=\theta=(f_0/f_1)^2=10^8$, and $q=\theta\epsilon=2\times 10^4$, so the conditions for (4) to be an approximate solution of (3) are satisfied. The values of the parameters for the acoustical warble tone do not always satisfy these conditions, and we have then to solve the equation

$$(1+2\epsilon\cos 2z)y''+\theta y=0 \qquad (6)$$

without using the above approximation. In succeeding sections it is shown how this may be accomplished to any desired degree of accuracy.

## 2. THEORETICAL CONSIDERATIONS

Since the coefficient of $y''$ in (6) Section 1 is single-valued and periodic in $z$, that of $y$ being

---

[1] J. R. Carson, "Notes on the theory of frequency modulation," Proc. I.R.E. 10, 62 (1922).

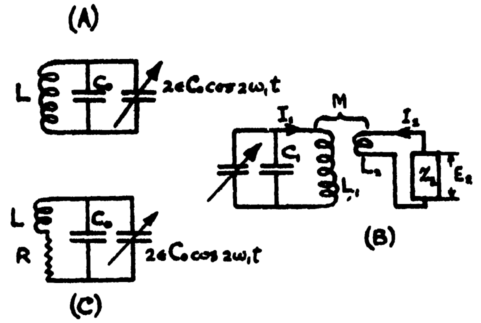


FIG. 1. (A) Schematic representation of loss-free oscillatory circuit with periodically variable capacitance $C=C_0(1+2\epsilon\cos 2\omega_1 t)$. (B) As at (A) but the circuit loosely coupled by $L_2$ to a second circuit. (C) As at (A) but the circuit having constant resistance $R$.

constant, Floquet's theory may be applied.[2] Thus the solutions are either stable, unstable, or neutral, according to the values of $\theta$ and $\epsilon$. For present purposes we shall consider stable solutions. Thus linearly independent formal solutions are[4]

$$y_1(z)=\sum_{-\infty}^{\infty} c_{2r}\cos(2r+\rho)z, \tag{1}$$

and

$$y_2(z)=\sum_{-\infty}^{\infty} c_{2r}\sin(2r+\rho)z, \tag{2}$$

provided the point $(\theta, \epsilon)$ lies in certain regions of the $(\theta, \epsilon)$ plane defined below, and $0<\rho<1$. For other regions of the plane, the solutions take the forms

$$y_1(z)=\sum_{-\infty}^{\infty} c_{2r+1}\cos(2r+1+\rho)z, \tag{3}$$

and

$$y_2(z)=\sum_{-\infty}^{\infty} c_{2r+1}\sin(2r+1+\rho)z. \tag{4}$$

These forms will be comprehended readily by reference to McLachlan,[3] Chapter 4, in which there are stability charts for Mathieu's equation (3) Section 1. In these the $(a, q)$ plane is mapped out into regions where the solutions are (a) stable, (b) unstable. A similar chart may be drawn for Eq. (6) Section 1, but in this case although $\theta$ is unlimited, $\epsilon$ is restricted to the strip $|\epsilon|<0.5$, as illustrated in Fig. 2. The values of $(\theta, \epsilon)$ in which we are interested, lie *within* the stable regions, being on the iso-$\rho$ curves. The full-line curves correspond to the characteristic curves for solutions of Mathieu's equation of positive integral order, and pertain to kindred solutions of (6) Section 1. They are designated $\alpha_m$, $\beta_m$, with $m=2n$, $2n+1$, or $2n+2$. When the point $(\theta, \epsilon)$ lies *between* $\alpha_{2n}$ and $\beta_{2n+1}$, the solutions (1), (2) are used, whereas for the regions *between* $\alpha_{2n+1}$ and $\beta_{2n+2}$, (3), (4) are used. If these solutions are normalized as shown in Section 12, they may then be regarded as *frequency modulation functions*, or f-m functions for brevity.

## 3. RECURRENCE RELATIONS FOR THE COEFFICIENTS IN (1)–(4) SECTION 2

Substituting either (1) or (2) into (6) Section 1, and equating the coefficient of $\cos(2r+\rho)z$, or that of $\sin(2r+\rho)z$ to zero, $r=-\infty$ to $+\infty$, we obtain the recurrence relation

$$[\theta-(2r+\rho)^2]c_{2r}-\epsilon[(2r+2+\rho)^2c_{2r+2} \\ +(2r-2+\rho)^2c_{2r-2}]=0. \tag{1}$$

The relation for (3), (4) Section 2 is

$$[\theta-(2r+1+\rho)^2]c_{2r+1}-\epsilon[(2r+3+\rho)^2c_{2r+3} \\ +(2r-1+\rho)^2c_{2r-1}]=0. \tag{2}$$

Both (1), (2) are linear difference equations of the second order, and each has two independent solutions, the character of which we shall now investigate. Dividing (1) throughout by $c_{2r}$ and writing $v_{2r}=c_{2r+2}/c_{2r}$ we get

$$(2r+2+\rho)^2v_{2r}+(2r-2+\rho)^2v_{2r-2}^{-1} \\ =[\theta-(2r+\rho)^2]\epsilon^{-1}, \tag{3}$$

or

$$v_{2r}+\left(\frac{2r-2+\rho}{2r+2+\rho}\right)^2 v_{2r-2}^{-1}=\left[\frac{\theta-(2r+\rho)^2}{(2r+2+\rho)^2}\right]\epsilon^{-1}. \tag{4}$$

With $\theta$ finite, $|\epsilon|<0.5$, $0<\rho<1$, when $r\to+\infty$ (4) may be written

$$v_{2r}+v_{2r-2}^{-1}\sim-\epsilon^{-1}. \tag{5}$$

Then as $r\to+\infty$, $|v_{2r}|$ is either $\gtrless 1$. For the convergence of (1), (2) Section 2, $|c_{2r}|\to 0$ as $r\to+\infty$, so this solution rests on the condition

---

[2] G. Floquet, "Sur les equations differentielles lineares," Ann. de l'Ecole Normale Superieure, 12, 47 (1883).

[3] N. W. McLachlan, *Theory and Application of Mathieu Functions* (Oxford Press, New York, 1947).

$|v_{2r}|<1$. The second solution $d_{2r}$ pertains to $|v_{2r}|>1$, which means that $|d_{2r}|\to\infty$ with $r$, so in the present instance it is inadmissible. Suppose that $|\epsilon|<0.1$, then to a first approximation we may write

$$v_{2r}\simeq-\epsilon. \tag{6}$$

Substituting from (6) into (5) leads to

$$|v_{2r-2}|\simeq|\epsilon/(1-\epsilon^2)|<1. \qquad (7)^*$$

As $r\to+\infty$, the ratio of consecutive coefficients tends to this approximate value. The same conclusion may be reached if $r\to-\infty$.

The formulas corresponding to (3), (4) Section 2 may be derived from those above by writing $(2r+1)$ for $2r$.

#### 4. MINIMUM RATIO OF CONSECUTIVE COEFFICIENTS

If the order of the solution (function) is $(m+p)$, where $m^2=4r_m^2$, it is convenient to regard $c_m=c_{2r_m}$ as the *central* coefficient in the solution. Then, if $s\geq r_m$, $|v_{2s}|=|c_{2s+2}/c_{2s}|<1$, and for $p\leq r_m$, $|1/v_{2p}|=|c_{2p-2}/c_{2p}|<1$, so $|v_{2p}|>1$. In (4) Section 3 let $r=-1$, and we get

$$v_{-2}+\left(\frac{4-\rho}{\rho}\right)^2 v_{-4}^{-1}=\left[\frac{\theta-(2-\rho)^2}{\rho^2\epsilon}\right]. \tag{1}$$

Now $0<\rho<1$, $|\epsilon|<0.5$, and if $\theta\gg1$, the r.h.s. of (1) may be relatively large. The second member of (1) is then negligible and, therefore,

$$v_{-2}\simeq[\theta-(2-\rho)^2]/\rho^2\epsilon. \tag{2}$$

For instance if $\theta=100$, $\epsilon=0.2$, $\rho=0.323$,

$$v_{-2}=c_0/c_{-2}\simeq4650/1, \tag{3}$$

so the minimum ratio of consecutive coefficients is 1/4650. As $r$ decreases below $-1$, the ratio increases and tends ultimately to the value at (7) Section 3. Hence, as $r$ increases above $r=s$, $|v_{2r}|$ decreases slowly to an ultimate value of $|\epsilon/(1-\epsilon^2)|$; and as $r$ decreases below $r=p$, $1/v_{2r}$ attains a minimum value of $\epsilon\rho^2/[\theta-(2-\rho)^2]$ at $r=-1$, but thereafter rises and ultimately approaches $-\epsilon/(1-\epsilon^2)$. Between $r=-1$ and $r=r_m$, the ratio is positive if $\epsilon>0$.

The formulas corresponding to (3), (4) Section 2, are

$$v_{-3}+\left(\frac{5-\rho}{1-\rho}\right)^2 v_{-5}^{-1}=[\theta-(3-\rho)^2]/(1-\rho)^2\epsilon, \tag{4}$$

and

$$v_{-3}\simeq[\theta-(3-\rho)^2]/(1-\rho)^2\epsilon, \tag{5}$$

this being the reciprocal of the minimum ratio. The behavior of the coefficients with variation in $r$ is that already described.

#### 5. CONVERGENCE OF (1)–(4) SECTION 2

In Section 4 it is shown that as $r\to\pm\infty$, the ratio of consecutive coefficients is less than unity, and that $|c_{2r}|\to0$. Now $\left|{\cos \atop \sin}(2r+\rho)z\right|\leq1$, if $z$ is real, so by a known theorem the series concerned are absolutely and uniformly convergent. When differentiated term by term repeatedly, the ratio of the moduli of consecutive coefficients is less than unity as $r\to\pm\infty$. Hence the resulting series have the same properties as above, as also have those obtained by integrating term by term, for a similar reason.

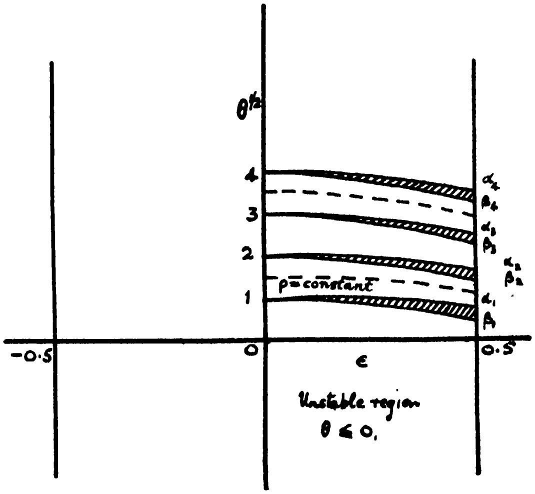


FIG. 2. Illustrating form of iso-$\rho$ chart for $(1+2\epsilon\cos2z)y''+\theta y=0$. $\alpha_1$, $\beta_1$, $\alpha_2$, $\beta_2$, $\cdots$ are the characteristic curves (solid) for solutions of integral order—not considered herein. The even-order characteristics are symmetrical about the $\theta^{\frac{1}{2}}$ axis, but those of odd order are asymmetrical although the chart is symmetrical. Iso-$\rho$ curves are indicated by broken lines. When $\theta>0$ and the point $(\theta^{\frac{1}{2}}, \epsilon)$ lies in an unshaded area, the solution is stable. If it lies in a shaded area between $\alpha_m$ and $\beta_m$, the complete solution is unstable and $\to\pm\infty$ as $z\to\pm\infty$.

---

* The sign of $v_{2r-2}$ is minus.

## 6. FORMULAS FOR CALCULATING TRIAL VALUES OF $\rho$

### a

Let

$$y = \cos\nu z + \epsilon c_1(z) + \epsilon^2 c_2(z) + \cdots, \qquad (1)$$

and

$$\theta = \nu^2 + \gamma_1\epsilon + \gamma_2\epsilon^2 + \cdots, \qquad (2)$$

where the $c(z)$ are continuous single-valued even functions of $z$; $\nu = (m+\rho)$ being real and positive. By substituting (1), (2) into (6) Section 1, taking the solution to be stable and the coefficient of $\cos\nu z$ unity, we may determine $\gamma_1$, $\gamma_2$, $\cdots$.[3] Then to a first approximation we find that

$$\theta \simeq \nu^2[1-(3\nu^2+4)\epsilon^2/2(\nu^2-1)]. \qquad (3)$$

If $\theta \gg 1$, and $\epsilon$ is small enough, (3) may be expressed in the form

$$\nu = (m+\rho)$$
$$\simeq [\theta(\theta-1)/\{\theta(1-\tfrac{3}{2}\epsilon^2)-(1+2\epsilon^2)\}]^{\frac{1}{2}}. \qquad (4)$$

Thus $m$ is the integral, and $\rho$ the fractional part of the r.h.s. of (4).

### b

We choose the form of differential equation at (2) Section 1 and expand $(1+2\epsilon\cos 2z)^{-1}$. Then

$$\begin{aligned}(1+2\epsilon\cos 2z)^{-1} = {} & (1+2\epsilon^2+6\epsilon^4+\cdots)\\ & -2\epsilon(1+3\epsilon^2+10\epsilon^4+\cdots)\cos 2z\\ & +2\epsilon^2(1+4\epsilon^2+\cdots)\cos 4z\\ & -2\epsilon^3(1+5\epsilon^2+\cdots)\cos 6z+\cdots. \qquad (1)\end{aligned}$$

If $\epsilon$ is such that terms in $\cos 4z$, $\cos 6z$, etc., may be neglected, we can write (2) Section 1 in the approximate form

$$y'' + (a-2q\cos 2z)y = 0, \qquad (2)$$

with

$$a = \theta(1+2\epsilon^2+6\epsilon^4+\cdots), \qquad (3)$$

and

$$q = \theta\epsilon(1+3\epsilon^2+10\epsilon^4+\cdots). \qquad (4)$$

Then from reference 3, page 83,

$$(m+\rho) \simeq [a-(a-1)q^2/\{2(a-1)^2-q^2\}]^{\frac{1}{2}}. \qquad (5)$$

The values of $\rho$ computed from either (4) **a**, or (5) are to be regarded as "trial" values, which may need to be improved during the process of computing the coefficients.

## 7. COMPUTATION OF THE COEFFICIENTS

Having obtained a "trial" $\rho$, the $c$'s are calculated by the method described in reference 3, Chapter 5. Unlike the $c$ therein, those to be computed for (1), (2) Section 2 do not decrease rapidly with increase in $r > r_m$, unless $\epsilon$ is very small. Moreover, it is necessary to introduce an approximate correction based upon (7) Section 3, when starting the computation.

## 8. EXAMPLE

### a

The following data pertain to an acoustical warble tone circuit: $f_0 = \omega_0/2\pi = 500$ c.p.s., $f_1 = \omega_1/2\pi = 50$ c.p.s., $\epsilon = 0.2$. $f_0$ is the central frequency which corresponds to the unmodulated condition $\epsilon = 0$. $\Delta f \simeq \epsilon f_0 = 100$ c.p.s., so the frequency range is approximately 400–600 c.p.s. The frequency of deviation or rate at which $f_0$ is modulated is $2f_1 = 100$ c.p.s., the modulation coefficient being $2\epsilon = 0.4$. As shown in Section 1, $|2\epsilon| < 1$. In the above data $f_1$ and $\epsilon$ are much greater than is usual in practice, where it is probable that $f_1 = 10$ c.p.s., and $\epsilon = 0.05$.

Using (4) **a** Section 6 with $\theta = (f_0/f_1)^2 = 100$ $\epsilon = 0.2$, yields

$$(m+\rho) = 10.322,$$

so

$$m = 10, \quad \text{and} \quad \rho = 0.322. \qquad (1)$$

Using (3)–(5) **b** Section 6, we find that $a = 109$, $q = 22.72$, giving

$$(m+\rho) = 10.323,$$

so

$$m = 10, \quad \text{and} \quad \rho = 0.323. \qquad (2)$$

Then $c_{10}$ is the central coefficient. The close agreement between the trial values of $\rho$ at (1), (2) should be noted, but this does not prove that they are accurate!

Taking one of the values of $\rho$, or the average, we calculate

$$\longrightarrow \quad c_2,\ c_4,\ c_6,\ c_8,\ c_{10},\ \breve{c}_{12},\ \breve{c}_{14},$$
$$\breve{c}_6,\ \breve{c}_8,\ c_{10},\ c_{12},\ c_{14},\ c_{16},\ c_{18}, \quad \longleftarrow$$

proceeding in the direction of the arrow in each case. Thus (a) we start at $c_2$ and work towards the central coefficient $c_{10}$, and (b) we start at $c_{18}$ and work towards $c_{10}$. In each case there is an overlap of two (or more if desired) coefficients, and if we have started far enough away from $c_{10}$

on either side, and the trial $\rho$ is sufficiently accurate, there ought to be adequate agreement between the pairs of coefficients $c_6$, $\bar{c}_6$; $c_8$, $\bar{c}_8$; etc. Failing this, re-computation using another value of $\rho$, and interpolation as in reference 3, Chapter 5, must be employed to attain the desired accuracy of the results.

### b. Procedure

First we shall show that, to the degree of accuracy contemplated in demonstrating the method of calculation (using a 10-inch slide rule) that $c_0$ may be neglected. By (3) Section 4, $c_{-2}=c_0/4650$, so with $r=0$, $\epsilon=0.2$, and $\rho=0.323$ in (1) Section 3, by neglecting $c_{-2}$ we obtain

$$\begin{aligned} c_0 &= \epsilon(2+\rho)^2c_2/(\theta-\rho^2) \\ &\simeq 0.0108c_2. \end{aligned} \tag{1}$$

With $r=1$ in (1) Section 3,

$$4.323^2c_4 = 5(100-2.323^2)c_2-0.323^2c_0, \tag{2}$$

$$= 472.8c_2-0.00113c_2, \tag{3}$$

by aid of (1). Thus the term $0.323^2c_0$ in (2) may be ignored, and we get

$$c_4 = 472.8c_2/4.323^2=25.1c_2. \tag{4}$$

$r=2$

$$6.323^2c_6 = 5(100-4.323^2)c_4-2.323^2c_2.$$

Substituting from (4) for $c_4$ we have

$$c_6 = 254.4c_2. \tag{5}$$

$r=3$

$$8.323^2c_8 = 5(100-6.323^2)c_6-4.323^2c_4,$$

so

$$c_8 = 1090c_2. \tag{6}$$

$r=4$

$$10.323^2c_{10} = 5(100-8.323^2)c_8-6.323^2c_6,$$

so

$$c_{10} = 1465c_2, \textit{ the central coefficient.} \tag{7}$$

It will be seen that owing to repetition, the squares of various numbers, e.g., $4.323^2$, need be obtained once only. We now continue and calculate the overlap coefficients $\bar{c}_{12}$, $\bar{c}_{14}$.

$r=5$

$$12.323^2\bar{c}_{12} = 5(100-10.323^2)c_{10}-8.323^2c_8,$$

so

$$12.323^2\bar{c}_{12} = 5(100-106.56)c_{10}-8.323^2c_8, \tag{8}$$

and, therefore,

$$\bar{c}_{12} = -814c_2. \tag{9}$$

This may be regarded as the "critical" point of the computation, because here $[100-(2r+0.323)^2]$ is smaller than for any other value of $r$. Consequently the influence of an inaccurate trial $\rho$ is greatest when $r=5$.

$r=6$

$$14.323^2\bar{c}_{14} = 5(100-12.323^2)\bar{c}_{12}-10.323^2c_{10},$$

so

$$\bar{c}_{14} = 270c_2. \tag{10}$$

### c

We now calculate the coefficients on the upper side of $c_{10}$, commencing with $c_{20}$. From (7) Section 3, the *limiting* value of the ratio of consecutive coefficients is approximately $-\epsilon/(1-\epsilon^2)=-0.2/0.96=-0.208$. But $|\bar{c}_{20}/c_{18}|>0.208$, so we shall tentatively assume that $c_{20}/c_{18}=-0.23$. Using (1) Section 3 with

$r=9$

$$16.323^2c_{16} = 5(100-18.323^2)c_{18}-20.323^2c_{20}. \tag{1}$$

Since

$$c_{20} = -0.23c_{18},$$

the last member of (1) is $95c_{18}$, so (1) yields

$$16.323^2c_{16}=-1180c_{18}+95c_{18}, \tag{2}$$

giving

$$c_{16}=-4.07c_{18}. \tag{3}$$

The second term on the r.h.s. of (2) is 8.05 percent of the first, so the error in taking $c_{20}/c_{18}=-0.23$ cannot exceed 1.85 percent, and in fact it is very much less. Enhanced accuracy would ensue by commencing with, say, $c_{26}$, but this is not justifiable for slide rule work. Continuing, we have

$r=8$ $$14.323^2c_{14}=5(100-16.323^2)c_{16}-18.232^2c_{18},$$

so $$c_{14}=14.8c_{18}. \tag{4}$$

$r=7$ $$12.323^2c_{12}=5(100-14.323^2)c_{11}-16.323^2c_{16},$$

so $$c_{12}=-44.5c_{18}. \tag{5}$$

$r=6$ $$10.323^2c_{10}=5(100-12.323^2)c_{12}-14.323^2c_{14},$$

so $$c_{10}=80c_{18}, \textit{ the central coefficient.} \tag{6}$$

$r=5$ $$8.323^2\bar{c}_8=5(100-10.323^2)c_{10}-12.323^2c_{12},$$

so $$\bar{c}_8=59.5_5c_{18} \text{ (critical point).} \tag{7}$$

$r=4$ $$6.323^2\bar{c}_6=5(100-8.323^2)c_8-10.323^2c_{10},$$

so $$\bar{c}_6=13.93c_{18}. \tag{8}$$

### d. Results in Terms of the Central Coefficient $c_{10}$

| | | Difference between overlap coefficients |
|---|---|---|
| $c_2=0.00068\,3c_{10}$ | | |
| $c_4=0.01715\ c_{10}$ | | |
| $c_6=0.1737\ c_{10}$ | $\bar{c}_6=0.1741\,c_{10}$ | $0.0004\,c_{10}$ |
| $c_8=0.745\ c_{10}$ | $\bar{c}_8=0.7444\,c_{10}$ | $-0.0006\,c_{10}$ |
| $c_{10}=1.00000\,0c_{10}$ | $c_{10}=1.0000\,c_{10}$ | $0.0000\,c_{10}$ |
| $\bar{c}_{12}=-0.556\ c_{10}$ | $c_{12}=-0.556\ c_{10}$ | $0.0000\,c_{10}$ |
| $\bar{c}_{14}=0.1845\ c_{10}$ | $c_{14}=0.185\ c_{10}$ | $0.0005\,c_{10}$ |
| | $c_{16}=-0.051\ c_{10}$ | |
| | $c_{18}=0.0125\,c_{10}$ | |

The agreement between $c_6$, $\bar{c}_6$; etc., is satisfactory for present purposes.

### e. Check on $\rho$

The agreement between the overlap coefficients in **d** does not warrant a check. When necessary, however, an improved value of $\rho$ may be found as follows: Put $r=5$ in (1) Section 3, and we get

$$[100-(10+\rho)^2]c_{10}-0.2[(12+\rho)^2c_{12}+(8+\rho)^2c_8]=0. \tag{1}$$

The computed values of $c_8$, $c_{12}$ are inserted in (1), which is then solved as a quadratic equation for $\rho$. Provided the difference between the trial $\rho$ and that obtained from (1) is a small fraction of the former, the new value of $\rho$ is likely to be the better value of the two. To the order of accuracy contemplated herein, the $\rho$ from (1) agrees with 0.323.

The complete computation may be checked by substituting (1) Section 2 into (6) Section 1 using the values of the $c$ in **d**, with $z=0$, or any other convenient value.

In **b** and **c** the computation has been set out deliberately in such a manner as to make the procedure easy to follow. For a systematic series of computations, the arrangement may be improved in a way which the professional computer will readily discern. Accuracy to as many decimal places as desired may be attained by well-known processes.

### 9. THE SOLUTION OF (1) SECTION 1

Writing $z=\omega_1 t$, $Q=y$ in (1), (2) Section 2, and using the $c$ in **d** Section 8, to the desired degree of approximation, we have

$$Q_1(t)\simeq\sum_1 c_{2r}\cos(2r+0.323)\omega_1 t, \quad (1)$$

and

$$Q_2(t)\simeq\sum_1^9 c_{2r}\sin(2r+0.323)\omega_1 t. \quad (2)$$

Hence the complete solution with two arbitrary constants is

$$Q(t)\simeq AQ_1(t)+BQ_2(t), \quad (3)$$

$$=C\sum_1^9 c_{2r}\cos[(2r+0.323)\omega_1 t-\phi], \quad (4)$$

where $C=(A^2+B^2)^{\frac{1}{2}}$, $\phi=\tan^{-1}(B/A)$. In practice either the circuit current or the e.m.f. induced in a coupled circuit is needed. The current is $dQ/dt$, while the e.m.f. is proportional to $dI/dt=d^2Q/dt^2$. Hence

$$dQ/dt\simeq -C\sum_1^9 (2r+0.323)\omega_1 c_{2r} \times\sin[(2r+0.323)\omega_1 t-\phi], \quad (5)$$

and

$$dI/dt\simeq -C\sum_1^9 (2r+0.323)^2\omega_1^2 c_{2r} \times\cos[(2r+0.323)\omega_1 t-\phi]. \quad (6)$$

The ratios of the coefficients in (4)–(6) in terms of the central coefficient, taken to be unity in each case, are set out in Table I.

### 10. SALIENT DIFFERENCES BETWEEN THE FORMS OF SOLUTIONS (4) OR (5) SECTION 1 AND (4) SECTION 9

If the numerical data satisfy conditions (a), (b) Section 1, the two forms of solution yield substantially the same results. But when this is not so, the procedure given in Section 8 must be used, and the following salient differences are exhibited:

TABLE I. Showing relative amplitudes of the side frequencies corresponding to $Q$, $I$, and $E$.

| $r$ | $Q(t)$ | $Q'(t)$ | $Q''(t)$ |
|---|---|---|---|
| 1 | 0.00068 3 | 0.00015 37 | 0.00003 47 |
| 2 | 0.01715 | 0.00719 | 0.00302 |
| 3 | 0.1737 | 0.1062 | 0.0649 |
| 4 | 0.745 | 0.6 | 0.483 |
| 5 | 1.00000 0 | 1.00000 0 | 1.00000 0 |
| 6 | −0.556 | −0.664 | −0.795 |
| 7 | 0.185 | 0.255 | 0.356 |
| 8 | −0.051 | −0.0806 | −0.127 |
| 9 | 0.0125 | 0.0222 | 0.0393 |

1. In (4) Section 1, the moduli of the coefficients equidistant from the central coefficient are equal, whereas in (4) Section 9 they are unequal, as will be seen from column 2 in Table I.

2. The coefficients in the series for $Q'(t)$, and $Q''(t)$ obtained by term differentiation of (4) Section 1, bear the same relation to each other as those in the series for $Q(t)$. The relative ratios for $Q'(t)$, and $Q''(t)$ obtained from (4) Section 9 differ from those for $Q(t)$, and the discrepancies are more marked the higher the derivative. Thus differentiation enhances the asymmetry about the central coefficient.

3. In (4) Section 1 the central frequency $f_0$, and the various side frequencies are affected to a trifling extent by capacitance variation, since $\epsilon$ is so small in a radio circuit. But owing to the relatively large value of $\epsilon$ in the warble tone circuit, in (4) Section 9, $f_0$ is increased by 3.23 percent to $1.0323f_0$, when $\epsilon=0.2$. Also each side frequency is increased by an amount dependent upon the value of $r$. For instance when $r=1$, the side frequency is $2.323f_1$ instead of $2f_1$, and when $r=9$, the corresponding values are $18.323f_1$ and $18f_1$. A common feature of the two forms of solution is that the side frequencies are spaced at integral multiples of $2f_1$ about the central frequency.

### 11. PROCEDURE WHEN $m$ IS LARGE

In Section 8 suppose that $f_1=10$ c.p.s., $\epsilon=0.05$, then $\theta=2500$, and $(m+\rho)\simeq\theta^{\frac{1}{2}}(1+3\epsilon^2/4)=50.375$. Thus $m=50$ and $\rho=0.375$. If the computation started at $c_2$ and were symmetrical about $c_{50}$, 49 coefficients would have to be calculated, thereby entailing considerable labor, especially if the trial $\rho$ were insufficiently accurate. When $\theta$ is large, this may be circumvented by calculating only the *significant* coefficients on each side of $c_m$. For a given $\theta$ the greater $\epsilon$ the larger is the number of significant coefficients. To determine the significant coefficient of lowest order, namely, $c_{2s}$ or $c_{2s+1}$, we use (1) or (2) Section 3, according as $m$ is even or odd. If $m$, $(m-2s)$ are large enough, the last member of (1) Section 3 may be neglected. Then if $s \gg 1$, we have

$$v_{2r}^{-1} = c_{2r}/c_{2r+2} \simeq 4(s+1)^2\epsilon/(\theta-4s^2). \qquad (1)$$

Provided adequate accuracy ensues when $v_{2r}^{-1}=h$, (1) yields the approximate formula

$$2s \simeq [h\theta/(h+\epsilon)]^{\frac{1}{2}}. \qquad (2)$$

With $\theta=2500$, $\epsilon=0.05$, $h=0.05$, (2) gives $2s=36$, so the number of significant coefficients on the lower side of $c_{50}$ is $(25-18)=7$. As in **b** Section 8 the value of $c_{34}$, the first coefficient omitted, may be calculated from $c_{34}/c_{36}=h$.

On the upper side of $c_m$, (1) Section 3 gives

$$v_{2r-2} = c_{2r}/c_{2r-2} \simeq 4(s-1)^2\epsilon/(\theta-4s^2), \qquad (3)$$

and (2) is usable here also. Hence including $c_m$, the total number of significant coefficients is $7+1+7=15$. If $h=0.1$ gave adequate accuracy, there would be eleven significant coefficients. Formula (2) is applicable when $m$ is odd, provided that $(2s+1)$ be written for $2s$. If $m$ is very large and $\epsilon$ very small (2) Section 1 may be written with adequate approximation

$$y''+\theta(1-2\epsilon\cos 2z)y=0, \qquad (4)$$

so the solution takes the form at (4) Section 1.

## 12. FREQUENCY MODULATION FUNCTIONS AND THEIR NORMALIZATION

Consider the coupled circuit shown schematically in Fig. 1B. If the coefficient of coupling is small, and the impedance of $Z_2$ much greater than that of the remainder of the circuit, the P.D. on $Z_2$ is nearly $E_2=MdI_1/dt$. Thus $I_2 \simeq E_2/Z_2$. If another circuit is coupled electromagnetically to $Z_2$, the P.D. across $Z_3$ will be approximately proportional to $d^2I_2/dt^2$, and so on. Thus the relative amplitudes of the side frequency P.D.S. in a particular circuit depend upon the number of circuits preceding it. In the $n$th circuit, the P.D. may be found by differentiating the series for $Q(t)$ in the first circuit $(n+1)$ times. This procedure is valid by virtue of uniform convergence as treated in Section 5. Accordingly it is proposed to base our definitions of the frequency modulation functions upon (1)–(4) Section 2, these being the forms of solution for the quantity $Q(t)$.

If in (1), (2) Section 2, $\rho=p/s$, a rational fraction in its lowest terms, the solutions are periodic in $z$ real, with period $2s\pi$, $s \geq 2$. Then if $K$ is a normalizing constant, we shall adopt the convention that[3]

$$\frac{K}{\pi s}\int_0^{2\pi s}\left[\sum c_{2r}\cos(2r+\rho)z\right]^2 dz$$
$$=\frac{K}{\pi s}\int_0^{2\pi s}\left[\sum c_{2r}\sin(2r+\rho)z\right]^2 dz$$
$$=K\sum_{-\infty}^{\infty} c_{2r}^2=1, \qquad (1)$$

so

$$K=\left[\sum_{\infty}^{\infty} c_{2r}^2\right]^{-\frac{1}{2}}. \qquad (2)$$

When $\rho$ is irrational, we use definition (2) above, i.e. we make $s\rightarrow+\infty$ in (1). Writing $E_{2r}^{(2n+\rho)}=Kc_{2r}$, we define the f-m functions of order $(2n+\rho)$ to be

$$\left.\begin{matrix} cf \\ sf \end{matrix}\right\}_{2n+\rho}(z,\epsilon)=\sum_{r=-\infty}^{\infty} E_{2r}^{(2n+\rho)} \begin{matrix}\cos\\ \sin\end{matrix}(2r+\rho)z, \qquad (3,4)$$

and those of order $(2n+1+\rho)$ to be

$$\left.\begin{matrix} cf \\ sf \end{matrix}\right\}_{2n+1+\rho}(z,\epsilon)=\sum_{r=-\infty} E_{2r+1}^{(2n+2+\rho)}$$
$$\times \begin{matrix}\cos\\ \sin\end{matrix}(2r+1+\rho)z. \qquad (5,6)^4$$

In this notation $c$, $s$, signify cos, sin, while $f$ indicates frequency modulation. When $\epsilon\rightarrow 0$, (3), (4)$\rightarrow\cos(2n+\rho)z$, $\sin(2n+\rho)z$, while (5), (6) $\rightarrow\cos(2n+1+\rho)z$, and $\sin(2n+1+\rho)z$, respectively. Thus all the $E$ tend to zero, except $E_{2n+1}^{(2n+\rho)}$, which tends to unity. Also when $\epsilon=0$, $\theta=(2n+\rho)^2$ or $(2n+1+\rho)^2$, as the case may be.

---

[4] For this case use (1), (2) with $(2r+1)$ for $2r$, and $K_1$ for $K$.

## 13. CHANGING THE SIGN OF $\epsilon$

This may be effected by writing $(\frac{1}{2}\pi - z)$ for $z$ in (6) Section 1. Making this change in (3)–(6) Section 12 yields solutions having the following properties: (a) they are neither odd nor even, (b) they do not reduce to the limiting forms in Section 12 when $\epsilon \rightarrow 0$, (c) they are unduly complicated. If we write $(-1)^r c_{2r}$ for $c_{2r}$, $(-1)^r c_{2r+1}$ for $c_{2r+1}$ in the respective recurrence relations (1), (2) Section 3, the sign of $\epsilon$ is changed. Using this substitution in (3)–(6) Section 12 with $-0.5 < \epsilon < 0$, we obtain the simple even and odd solutions (f-m functions)

$$\left.\begin{matrix} cf \\ sf \end{matrix}\right\}_{2n+p}(z, -\epsilon) = (-1)^n \sum_{-\infty}^{\infty} (-1)^r E_{2r}^{(2n+p)} \begin{matrix} \cos \\ \sin \end{matrix} (2r+p)z, \quad (1, 2)$$

and

$$\left.\begin{matrix} cf \\ sf \end{matrix}\right\}_{2n+1+p}(z, -\epsilon) = (-1)^n \sum_{-\infty}^{\infty} (-1)^r E_{2r+1}^{(2n+1+p)} \begin{matrix} \cos \\ \sin \end{matrix} (2r+1+p)z. \quad (3, 4)$$

The purpose of the multiplier $(-1)^n$ is to insure a conventional positive sign when $\epsilon = 0$, and the functions reduce to $+\cos(m+p)z$, and $+\sin(m+p)z$, respectively, with $m = 2n$ or $2n+1$.

## 14. OSCILLATORY CIRCUIT HAVING CONSTANT RESISTANCE BUT PERIODICALLY VARIABLE CAPACITANCE

This is shown schematically in Fig. 1C, the differential equation being

$$d^2Q/dt^2 + (R/L)dQ/dt + Q/LC_0(1 + 2\epsilon \cos 2\omega_1 t) = 0. \quad (1)$$

Using the substitutions given in Section 1, and putting $R/2L = \kappa$, we get

$$(1 + 2\epsilon \cos 2z)(y'' + 2\kappa y') + \theta y = 0. \quad (2)$$

Writing $y = e^{-\kappa z} u(z)$, (2) is transformed to

$$(1 + 2\epsilon \cos 2z)u'' + (\theta_1 - 2\epsilon_1 \cos 2z)u = 0, \quad (3)$$

with $\theta_1 = (\theta - \kappa^2)$, and $\epsilon_1 = \epsilon\kappa^2$. The formal solutions of (3) are of the type given at (1)–(4) Section 3. Substituting (1) Section 2 into (3) above, and equating the coefficient of $\cos(2r+p)z$ to zero for $r = -\infty$ to $+\infty$ yields the recurrence relation

$$[\theta_1 - (2r+p)^2]c_{2r} - [\epsilon(2r-2+p)^2 + \epsilon_1]c_{2r-2} - [\epsilon(2r+2+p)^2 + \epsilon_1]c_{2r+2} = 0. \quad (4)$$

The recurrence relation corresponding to (3), (4) Section 2, is (4) with $(2r+1)$ for $2r$. By analysis akin to that in Section 6, formulas for calculating $m$ and $p$ may be derived. Then the procedure follows on the lines of that in Section 8 *et seq.* Approximate solutions of (1) Section 1 and (1) above are given in reference 3, pages 280 and 283, the form being quite different from that herein.

# The Sensitized Luminescence of Manganese-Activated Calcite

James H. Schulman, Lyle W. Evans, and Robert J. Ginther
*Naval Research Laboratory, Washington, D. C.*

AND

K. J. Murata
*Geological Survey, U. S. Department of the Interior, Washington, D. C.**

(Received January 20, 1947)

Synthetic manganese-activated calcites are shown to be practically inert to ultraviolet excitation in the range 2000–3500A, while they are luminescent under cathode-ray excitation. The incorporation of small amounts of an auxiliary impurity along with the manganese produces the strong response to ultraviolet radiation hitherto ascribed to $CaCO_3$:Mn itself. Three such impurities have been studied: lead, thallium, and cerium. The first two induce excitation in the neighborhood of the mercury resonance line, while the cerium introduces a response principally to longer wave ultraviolet. The strong response to 2537A excitation shown by some natural calcites is likewise found to be due to the presence of lead along with the manganese, rather than to the manganese alone. The data do not warrant ascribing the longer wave-length ultraviolet-excited luminescence of all natural calcites to the action of an auxiliary impurity. The essential identity of the cathode-ray excited luminescence spectra of $CaCO_3$:Mn, $CaCO_3$:(Pb+Mn), $CaCO_3$:(Tl+Mn), and $CaCO_3$:(Ce+Mn) with the 2537A-excited spectra of the latter three is evidence that the luminescent center in all cases is the manganese ion or the $MnO_6$ group. It is shown that a "cascade" mechanism for the action of the auxiliary impurities, lead, thallium, and cerium, is incorrect; and that the phenomenon must be considered as a case of sensitized luminescence. Owing to the nature of cathode-ray excitation, the manganese activator can be excited by this agent even in the absence of a second impurity. For optical excitation, however, an absorption band for the ultraviolet must be established by building into the $CaCO_3$:Mn a second impurity or "sensitizer."

THE rose to orange-red luminescence of natural calcite has been studied by many investigators[1–5] and has been ascribed to the presence of manganese as an activator. The preparation of a synthetic manganese-activated calcite phosphor was described by Fonda,[6] who precipitated $CaCO_3$:Mn from a $CaCl_2$-$MnCl_2$ solution by means of $(NH_4)_2CO_3$. This material fluoresced with a rose color under 2537A excitation, the best preparation having 30 percent of the brightness of calcite from Franklin, New Jersey. Our attempts to duplicate this preparation yielded material much inferior in brightness to Fonda's product, judging by comparison with the best sample of calcite that we could obtain from the above locality.

It has recently been shown by Murata and Smith[7] that manganese-activated sodium chloride is not excited by 2537A radiation unless lead is also present as a "co-activator" in the phosphor. This observation, together with our inability to duplicate Fonda's work, suggested that an unsuspected impurity—possibly lead—was at least partially responsible for the luminescence of his preparations, and that the amount of this impurity was smaller in our work. This was found to be the case. When the reagents used to prepare the calcite were first purified before following Fonda's procedure for precipitation of the phosphor,** the product was almost completely inert to both 2537A*** and "3650A"*** excitation.

---

* Data on natural calcites published by permission of the Director, Geological Survey, U. S. Department of Interior.

[1] E. L. Nichols, H. L. Howes, and D. T. Wilber, Phys. Rev. **12**, 351 (1918).

[2] T. Tanaka, J. Opt. Soc. Am. **8**, 287 (1924).

[3] E. L. Nichols, H. L. Howes, and D. T. Wilber, Carnegie Inst. Rep. **384**, 367 (1928).

[4] W. L. Brown, Univ. Toronto Studies, Geol. Series **36**, 45 (1934).

[5] P. Pringsheim, *Fluoreszenz and Phosphoreszenz* (Julius Springer, Berlin, 1928), p. 312.

[6] G. R. Fonda, J. Phys. Chem. **44**, 435 (1939).

[7] K. J. Murata, and R. L. Smith, Am. Mineral **31**, 527 (1946).

** Preparation of the phosphors was carried out by Fonda's procedure, precipitation of the $CaCO_3$ from a $CaCl_2-MnCl_2$ solution at 70° by means of 50 percent excess $(NH_4)_2CO_3$ solution, followed by aging of the precipitate in its supernatant liquor for 10 minutes at the boiling point. 100 ml volumes of $CaCl_2-MnCl_2$ solution

Luminescence could be produced in the final calcite product by adding to the calcium chloride solution an auxiliary component—either a lead, thallium, or cerium salt—along with the manganese chloride prior to precipitation of the phosphor. The addition of lead or thallous chlorides as auxiliary impurities, giving $CaCO_3:(Pb+Mn)$ and $CaCO_3:(Tl+Mn)$, respectively, induced strong luminescence under 2537A excitation, but only a minor increase in the "3650A" excited luminescence. The color of the luminescence was rose, in samples of low brightness, to orange-red in samples of high brightness.† The brightest preparations of $CaCO_3:(Pb+Mn)$ were 90 percent as bright, under 2537A excitation, as an outstandingly efficient specimen of calcite from Langban, Sweden.

The addition of cerium to the $CaCl_2$-$MnCl_2$ solution (as cerous ammonium nitrate) prior to the precipitation gave a product that luminesced both under 2537A and "3650A" excitation, with a rose color in both cases. Enhancement of the response to the latter excitation was greater with cerium than with lead or thallium as the auxiliary impurity, cerium-bearing samples being much brighter than the $CaCO_3:Mn$.

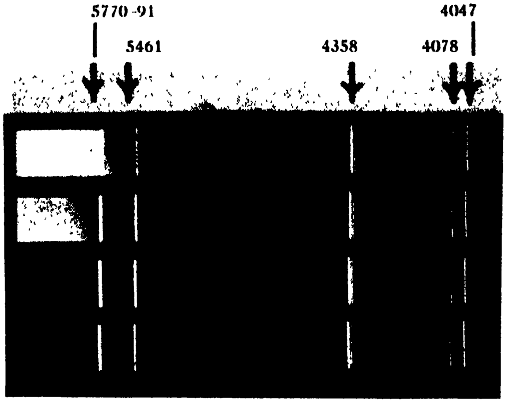


FIG. 1. Luminescence of synthetic manganese-activated calcites under 2537A excitation. Gaertner L231 spectrometer with camera attachment. Type II-F plate. Exposure times: 25 minutes. Exciting source: "Mineralite" lamp without filter.

A. Cadmium borate: Mn
B. Calcite: (Pb+Mn)
C. Calcite: (Tl+Mn)
D. Calcite: (Ce+Mn).

## LUMINESCENCE UNDER OPTICAL EXCITATION

The 2537A-excited luminescence spectra of synthetic calcites containing lead, thallium, and cerium additions are shown in Figs. 1 and 2. Included also in Fig. 1 is the 2537A-excited luminescence spectrum of a sample of manganese-activated cadmium borate phosphor having a quantum efficiency of 75–80 percent under this excitation. The spectra given by the three calcite preparations are essentially identical, starting at about 5780A and extending beyond the limits of sensitivity of the plate used.

Measurements of the brightness of a series of preparations of $CaCO_3:(Pb+Mn)$, $CaCO_3:(Tl+Mn)$, and $CaCO_3:(Ce+Mn)$ are given in

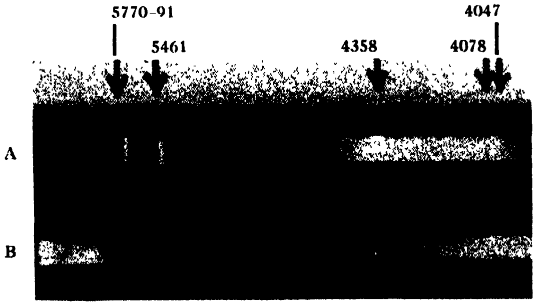


FIG. 2. Luminescence of synthetic calcite: (Ce+Mn) under 2537A excitation. Gaertner L231 spectrometer with camera attachment. Type II-F plate.

A. "Mineralite" lamp, no filter, 1 minute exposure.
B. Calcite: (Ce+Mn) excited by "Mineralite" through Corning 9863 filter. Exposure time: 16 hours. Luminescence between 4000–5000A caused by fused silica of lamp.

---

were used in all the experiments. The concentrations of our solutions were: $CaCl_2$, 0.162 molar; $(NH_4)_2CO_3$, 0.25 molar. These concentrations are twice as high as those employed by Fonda.

Purification of the $CaCl_2$ solution was accomplished by partial (33 percent) precipitation with $(NH_4)_2CO_3$ at 70°. Judging by the inertness of the final product, a greater degree of purification was effected by this procedure than by treatment of the $CaCl_2$ solution with $H_2S$. The $(NH_4)_2CO_3$ solution was purified by partial (33 percent) precipitation with $CaCl_2$ solution at 70°. It was not found necessary to purify the $MnCl_2$ solution, but this is desirable in general. Purification of the $MnCl_2$ may be effected by the same procedure used for purification of the $CaCl_2$.

*** 2537A excitation was provided by a low pressure mercury vapor lamp ("Mineralite" or "germicidal" lamp) equipped with a Corning No. 9863 filter. "3650A" excitation was provided by a 360BL lamp with the same filter, giving a broad band in the near ultraviolet (3300–4200A) with a peak near 3650A.

† Spectroscopic comparison of the luminescence of several preparations showed that the rose-colored fluorescence of the weaker samples is due to the superposition of visible (violet and blue) light from the exciting source upon the light emitted by the sample; during the brief afterglow following cessation of excitation the orange-red color of the luminescence is seen, since visible light from the lamp is no longer superimposed. The reflection of light from the exciting source becomes less important the brighter the phosphor, and the true orange-red luminescence color is therefore observed with these brighter samples even during excitation.

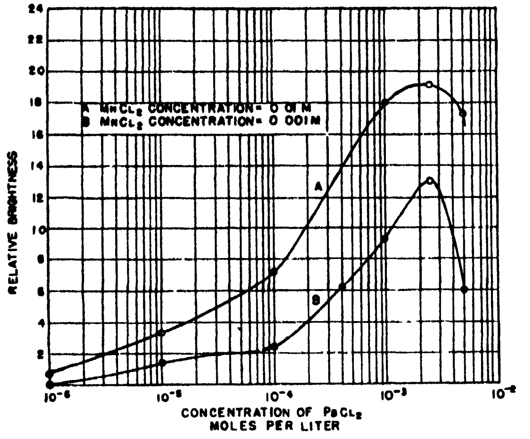


FIG. 3. Effect of lead concentration on brightness of synthetic calcites.

Table I, relative to the brightness of the above mentioned cadmium borate phosphor. In these measurements, circular plaques of the samples, one inch in diameter, were illuminated by a "Mineralite" lamp with a Corning 9863 filter and viewed by a barrier layer photo-cell through a red filter cutting off visible light below 5500A. A plaque of unactivated CaCO$_3$ was used to obtain the "dark response" of the system, which was due to the small amount of visible radiation from the lamp passed by both filters. The small correction obtained in this way was applied to all the readings.

The data on CaCO$_3$:(Pb+Mn) are further graphically illustrated in Figs. 3 and 4. An x-ray diffraction study of all the samples indicated that a number of them were mixtures of calcite and aragonite, instead of pure calcite. Those preparations in which aragonite was found are indicated by ● in the figures, while the pure calcite samples are indicated by ○.[††]

## LUMINESCENCE UNDER CATHODE-RAY AND SPARK-COIL EXCITATION

Visual observations were made of the cathode-ray excited luminescence of a group of synthetic calcite phosphors of the same manganese concentration, including a sample containing manganese alone, and samples containing manganese plus one of each of the auxiliary impurities: lead, thallium, and cerium. The phosphor samples were mounted side by side in $\frac{1}{8}''$ diameter$\times\frac{1}{16}''$ deep wells drilled into a brass plate, which was placed inside a demountable cathode-

TABLE I. Brightness of CaCO$_3$:(Pb+Mn), CaCO$_3$:(Tl+Mn), and CaCO$_3$:(Ce+Mn) phosphors under 2537A excitation.

| Molality of solution** with respect to | | Relative brightness |
|---|---|---|
| Pb$^{++}$ | Mn$^{++}$ | |
| $5\times10^{-3}$ | $10^{-2}$ | 17.2 |
| $5\times10^{-3}$ | $10^{-3}$ | 6.1 |
| $2.5\times10^{-3}$ | $10^{-2}$ | 19.2 |
| $2.5\times10^{-3}$ | $10^{-3}$ | 13.0 |
| $10^{-3}$ | $5\times10^{-2}$ | 0.26 |
| $10^{-3}$ | $3\times10^{-2}$ | 2.32 |
| $10^{-3}$ | $10^{-2}$ | 18.0 |
| $10^{-3}$ | $5\times10^{-3}$ | 19.2 |
| $10^{-3}$ | $3\times10^{-3}$ | 18.2 |
| $10^{-3}$ | $10^{-3}$ | 9.3 |
| $10^{-3}$ | $10^{-4}$ | Too weak to measure |
| $4\times10^{-4}$ | $10^{-1}$ | 6.3 |
| $10^{-4}$ | $10^{-2}$ | 7.2 |
| $10^{-4}$ | $10^{-3}$ | 2.44 |
| $10^{-4}$ | $10^{-4}$ | Too weak to measure |
| $10^{-5}$ | $10^{-2}$ | 3.3 |
| $10^{-5}$ | $10^{-3}$ | 1.52 |
| $10^{-5}$ | $10^{-4}$ | Too weak to measure |
| $10^{-6}$ | $10^{-2}$ | 0.76 |
| $10^{-6}$ | $10^{-3}$ | Too weak to measure |
| $10^{-6}$ | $10^{-4}$ | 0. |
| Tl$^{+}$ | Mn$^{++}$ | |
| $10^{-3}$ | $10^{-2}$ | 5.8 |
| $10^{-3}$ | $10^{-3}$ | Too weak to measure |
| $10^{-3}$ | $10^{-4}$ | 0. |
| $10^{-4}$ | $10^{-2}$ | 1.55 |
| $10^{-4}$ | $10^{-3}$ | Too weak to measure |
| $10^{-4}$ | $10^{-4}$ | 0. |
| Ce$^{+++}$ | Mn$^{++}$ | |
| $10^{-3}$ | $10^{-2}$ | 0.64*(*b*) |
| $10^{-3}$ | $10^{-3}$ | 0.71*(*a*) |
| $10^{-3}$ | $10^{-4}$ | 0. |
| $10^{-4}$ | $10^{-2}$ | 0.79*(*c*) |
| $10^{-4}$ | $10^{-3}$ | Too weak to measure |
| $10^{-4}$ | $10^{-4}$ | 0. |
| Calcite, Franklin, New Jersey | | 12.4 |
| Calcite, Langban, Sweden | | 21.3 |
| Cadmium borate: manganese | | 100. |

* Brightness under "3650A" excitation: (*a*) > (*b*) > (*c*).

** Polarographic and colorimetric analyses of several of the above samples showed that the mole ratios of Mn/Ca, Pb/Ca, and Ce/Ca in the solid were substantially the same as in the solution prior to precipitation, while in the case of thallium the Tl/Ca ratio in the phosphor was only one seventh of its value in the solution prior to precipitation. These results are in agreement with what is expected from elementary chemical considerations concerning the solubilities of the various carbonates, basic carbonates, and hydroxides involved.

---

[††] The preparation of the pure calcite form of CaCO$_3$, CaCO$_3$:Pb, CaCO$_3$:Tl, and CaCO$_3$:Ce by the hot precipitation, recommended by Fonda and used by us in most of the work, was found to be impossible, aragonite or, at best, mixtures of calcite and aragonite being formed instead. Apparently manganese has to be co-precipitated with the calcium carbonate in order to generate the calcite phase, although even in this case the formation of aragonite is not always prevented. Manganese-free calcite, with or without Pb, Tl, or Ce impurity, could be prepared only by low temperature precipitation. To obtain these materials a saturated calcium bicarbonate solution was made by bubbling carbon dioxide through a calcium carbonate suspension, filtering off the excess solid, and adding PbCl$_2$, TlCl, or CeCl$_3$ solution to the filtrate. From such solutions calcite was deposited at room temperature after a day or two.

ray tube. There was no perceptible color difference between the luminescence of $CaCO_3$:Mn, $CaCO_3$:(Pb+Mn), $CaCO_3$:(Tl+Mn) and $CaCO_3$:(Ce+Mn) when they were simultaneously bombarded with cathode rays (accelerating voltage, 800 volts; current density approximately 100 $\mu$amps/cm$^2$). The close proximity of the samples would tend to make any differences between them readily detectable.

Since the design of the demountable cathode-ray tube used in making these visual observations was not suited for use as a spectrographic source unit, the luminescence spectra of the above phosphors under spark-coil excitation, in air at a pressure of $10^{-2}$ mm Hg, were photographed instead (Fig. 5). It has been our observation that this type of excitation and cathode-ray excitation give the same luminescence spectrum with all the standard sulfide, silicate, fluoride, and tungstate phosphors. The visual observations and the spectrograms of Fig. 5 show the essential identity of the luminescence under corpuscular excitation of manganese-activated calcites with or without auxiliary impurities.

## NATURAL CALCITES

In order to determine whether or not the luminescence of natural manganese-bearing calcites is due to an auxiliary impurity, as with our synthetic samples, preliminary qualitative spectrographic analyses were made of brightly orange-red fluorescing (2537A excitation) natural calcites from the following localities: (1) Franklin, New Jersey; (2) Langban, Sweden; (3) Mount Spurr, Alaska Peninsula, Alaska; (4) Sunset Mine, Superior, Arizona; and (5) Guanajuato, Mexico. All of them were found to contain both manganese and lead. The manganese content of the samples ranged from tenths of a percent to several percent; and the lead from hundredths of a percent to over a percent. Neither thallium nor cerium could be detected in any of the samples, which therefore are samples of lead-induced luminescence. It follows that the "3650A" excited luminescence of these natural calcites is induced by some other impurity than those put into our synthetic samples, or is due to the manganese alone.

Chemical analysis of the outstandingly bright Langban calcite showed that it contained 1.3 weight percent of lead and 3.0 weight percent of manganese, which is the order of magnitude of both impurities present in our synthetic samples.

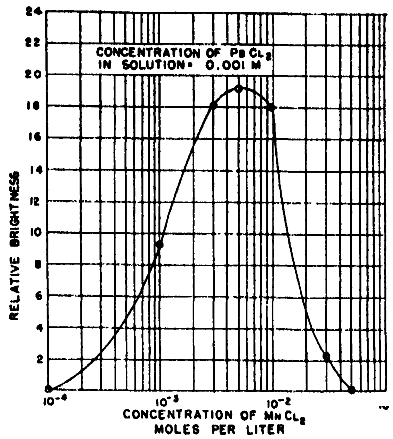


FIG. 4. Effect of manganese concentration on brightness of synthetic calcites.

The well-formed crystals from Guanajuato, Mexico, showed a sharply-zoned fluorescence that was restricted to the outermost millimeter of the crystals. Spectrographic analysis of the fluorescent and non-fluorescent portions showed that the manganese content was approximately 0.1 percent in both. The fluorescent portion, however, also contained several tenths of a

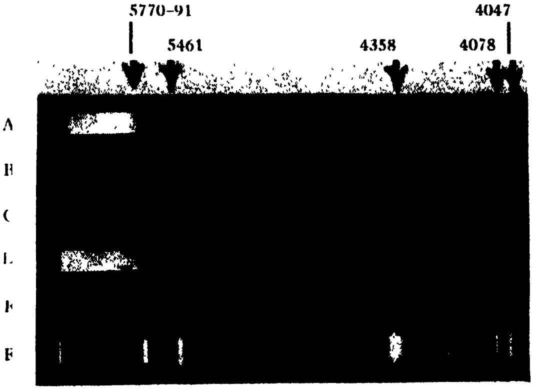


FIG. 5. Luminescence of manganese-activated calcites under spark-coil excitation. Gaertner L231 spectrometer with camera attachment. Type II-F plate. Exposure times: 5 minutes. Density differences between samples not significant.

A. Cadmium borate:Mn
B. Calcite:Mn
C. Calcite:(Pb+Mn)
D. Calcite:(Tl+Mn)
E. Calcite:(Ce+Mn)
F. "Mineralite" lamp, no filter, 1 minute exposure (for wave-length calibration).

percent of lead, whereas the non-fluorescent portion had no detectable lead.

The preliminary study of natural calcites thus confirms the results obtained with our synthetic samples, and demonstrates that the luminescence of natural calcites under excitation in the region of mercury resonance radiation is due—not to manganese alone, as has hitherto been believed—but to the simultaneous presence of manganese and lead in the crystal.

## DISCUSSION

The lack of fluorescence response of $CaCO_3$:Mn to optical excitation by ultraviolet radiation in the neighborhood of the mercury resonance line, and the development of such a response by the incorporation of auxiliary impurities such as lead, thallium, or cerium, may be considered from three different viewpoints. First, it may be postulated that the manganese activator alone in the calcite structure does not constitute a luminescent center, but that luminescence emission is possible only from a more complex center containing lead, thallium, or cerium in addition to the manganese. The auxiliary impurities and the manganese are accordingly interpreted as "co-activators."[7] This interpretation is ruled out by the fact that all samples of manganese-activated calcite, whether or not they luminesced under ultraviolet excitation, responded to cathode-ray and spark-coil excitation with the same luminescence emission spectrum. The cathode-ray excited luminescence is essentially identical with the optically-excited luminescence of the preparations responsive to the latter type of excitation. This luminescence is of the typical orange-red color associated with manganese in sixfold coordination with oxygen,[8, 9] as would be expected if $Mn^{++}$ replaced $Ca^{++}$ substitutionally in calcite. This shows that the $Mn^{++}$ ions or the $MnO_6$ groups existing in the crystal constitute in themselves luminescent centers, independent of the presence or absence of an auxiliary impurity and of the nature of the impurity when present.

A second interpretation is that we are dealing here with a "cascade" process in the case of ultraviolet excitation of the phosphors containing an auxiliary impurity along with the manganese. According to this simple picture, the manganese and the auxiliary impurity need not be considered as forming a manganese ion-impurity ion complex, but act as two separate luminescent centers. The auxiliary impurity absorbs the exciting "short" ultraviolet radiation and emits as fluorescence longer wave ultraviolet or even visible radiation capable of exciting the manganese activator to its typical luminescence. This interpretation assumes, then, a greater or less degree of coincidence between the position of the manganese absorption band in $CaCO_3$:Mn and the emission bands of $CaCO_3$:Pb, $CaCO_3$:Tl, and $CaCO_3$:Ce. This interpretation is contradicted by simple experiments in which intimate mechanical mixtures of $CaCO_3$:Mn with $CaCO_3$:Pb, $CaCO_3$:Tl, and $CaCO_3$:Ce, respectively, were examined under 2537A excitation. In no case was luminescence observed.

The third interpretation of our results is that the $CaCO_3$:(Pb+Mn), $CaCO_3$:(Tl+Mn), and $CaCO_3$:(Ce+Mn) phosphors are cases of "sensitized luminescence," such as those first described by Rothschild[10] in the group of rare-earth activated alkaline earth sulfide phosphors. In this model it is assumed that (a) $CaCO_3$:Pb, $CaCO_3$:Tl, and $CaCO_3$:Ce have an absorption band in the short wave-length ultraviolet region in question; (b) $CaCO_3$:Mn does not have an absorption band in this region of the ultraviolet, but may have absorption bands in another spectral range; (c) the simultaneous incorporation of manganese and lead, thallium, or cerium in calcite leaves conditions (a) and (b) substantially unchanged; (d) energy absorbed in the lead, thallium, or cerium-induced absorption bands is transferred to the manganese by a process analagous to collisions of the second kind rather than by a radiative ("cascade") process; and (e) the luminescence emission is due to the $Mn^{++}$ ion alone and is independent in spectral characteristics of how the energy is imparted to this ion.

The fundamental lattice absorption of pure calcite lies at wave-lengths shorter than about

---

[8] S. H. Linwood, and W. A. Weyl, J. Opt. Soc. Am. **32**, 443 (1942).

[9] J. H. Schulman, J. App. Phys. **17**, 902 (1946).

[10] S. Rothschild, Physik Zeits. **35**, 557 (1934); *ibid.* **37**, 757 (1936).

2250A.[11] The introduction of absorption bands at longer ultraviolet wave-lengths than the fundamental crystal absorption by the incorporation of lead, thallium, silver, and other heavy metal impurities in the alkali halides is well-known,[12] and a similar effect is to be expected in crystals with anions other than the halogens, although the position of the absorption peak produced by a given cation impurity will differ with the nature of the host crystal. The absorption produced by the impurity ion may be due to the excitation of this ion itself, as proposed by Seitz[13] for the alkali halide-thallium phosphors; or to the higher ionization potential of the substituting cation compared to that of the cation of the crystal itself[14]; or to a greater polarizing power of the impurity ion compared to the cation of the host crystal. In our manganese-activated calcites, the lead, thallium, and cerium impurities would act in this way as "absorbers" or "sensitizers," and this function would be performed whether the "sensitizers" themselves are capable of emitting luminescent radiation of their own or not. The manganese activator and the lead, thallium, or cerium "sensitizer" would have to be near neighbors in the calcite structure for the energy to be transferred from the sensitizer to the manganese, but this requirement is different from the concept of a $Mn^{++}-Pb^{++}$ (or $-Tl^{+}$ or $-Ce^{+++}$) complex as the luminescent center.

According to this picture of sensitized luminescence, no sensitizer should be required for the appearance of the manganese luminescence under cathode-ray bombardment. Under this type of excitation, energy is absorbed primarily by the atoms of the host crystal, the absorbed energy then being capable of transmission over large distances through the lattice to the site of the manganese activator by means of excitons or by the movement of electrons and positive holes through the crystal.[15]

The validity of the above interpretation of sensitized luminescence is supported by a more detailed study of the excitation and absorption (reflection method) of manganese-activated calcite phosphors being made by Mr. Elias Burstein and Mrs. Margaret White, Crystal Section, N.R.L., whom we thank for permission to present a summary of their preliminary results. Data on the excitation and reflection spectra of the lead, thallium, and cerium sensitized phosphors, as well as of calcites containing lead, thallium, cerium, and manganese separately, and of pure calcite are given in Figs. 6 and 7.

For the excitation studies, plaques of phosphor powder were irradiated with ultraviolet from a Beckman hydrogen discharge lamp, monochromatized by a Beckman quartz spectrophotometer, and the irradiated surface viewed by an RCA 1P21 multiplier photo-tube through a Corning 2418 filter. The design of the apparatus allowed the continuous automatic plotting of the photo-cell response *vs.* excitation wave-length in about 15 minutes over the range of 2000 to 3500A. The data are presented in Fig. 6 as cell response *vs.* wave-length instead of as a true excitation spectrum, since no means was provided for keeping constant ultraviolet energy incident on the phosphor at all wave-lengths or for measuring the intensity of the incident ultraviolet as a function of wave-length. Since the hydrogen lamp employed gives a continuous spectrum without sharp changes in output with wave-length, and the transmission of the monochroma-

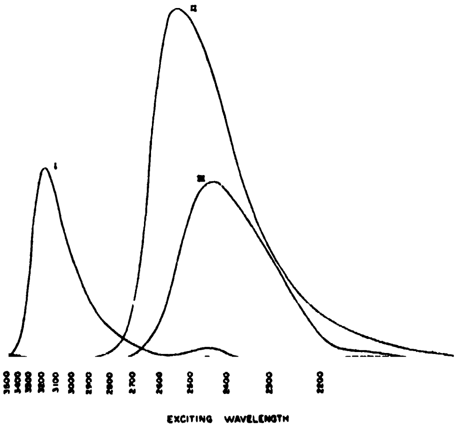


FIG. 6. Ultraviolet excitation spectra of synthetic manganese-activated calcites.

I. Calcite: (Ce+Mn)
II. Calcite: (Pb+Mn)
III. Calcite: (Tl+Mn)
Dotted curve - calcite: Mn.

---

[11] International Critical Tables, Vol. VII, pp. 270–273.

[12] R. Hilsch, Physik. Zeits. 38, 1031 (1927).

[13] F. Seitz, J. Chem. Phys. 6, 150 (1938).

[14] N. F. Mott, and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Oxford University Press, 1940), p. 102.

[15] See reference 14, p. 207.

tor likewise varies continuously, the data suffice to locate qualitatively, at least, the positions of excitation bands. Pure natural calcite (curve not shown in Fig. 6) gave no response over the whole range. Synthetic calcite, activated with manganese alone (dotted curve, Fig. 6) showed only a very faint luminescence under excitation near 2150A and again near 2450A. It is not yet known whether this very slight response is due to manganese alone or to traces of a sensitizing impurity which our purification method failed to remove. With $CaCO_3$:(Ce+Mn) (curve I) the 2450A excitation peak is somewhat higher, but a much more important effect is the far stronger excitation band created by the cerium at longer wave-lengths, peaking at about 3170A. $CaCO_3$:(Tl+Mn) (curve III) has a strong excitation region peaking at 2435A, while $CaCO_3$:(Pb+Mn) (curve II) shows a very intense excitation region with a peak at 2540A. This last band coincides almost exactly with the wave-length of mercury resonance radiation, which probably accounts for the much higher efficiency of luminescence of the $CaCO_3$:(Pb+Mn) phosphors under this excitation compared to that of $CaCO_3$:(Tl+Mn) and $CaCO_3$:(Ce+Mn) (*cf.* Table I).

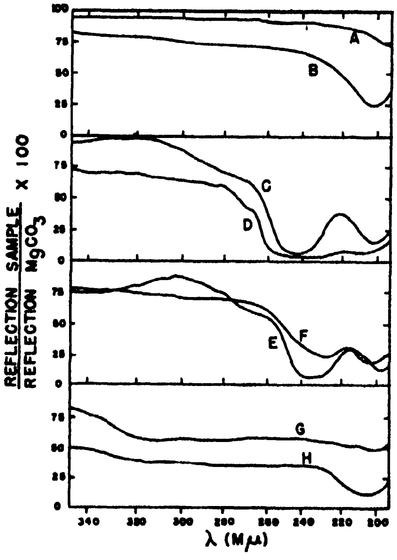


FIG. 7. Reflection spectra of calcite preparations.

A. Unactivated calcite (optical quality, non-luminescent natural specimen)
B. Calcite:Mn
C. Calcite:Pb
D. Calcite:(Pb+Mn)
E. Calcite:Tl
F. Calcite:(Tl+Mn)
G. Calcite:Ce
H. Calcite:(Ce+Mn).

In Fig. 7 are given the results of reflection measurements on the above materials. For these measurements the samples were illuminated with a Hanovia hydrogen-discharge lamp, and the reflected light analyzed by a Gaertner quartz monochromator, using a 1P28 multiplier phototube as a receiver. A plaque of magnesium carbonate[16] was used as a standard and the photocell response to the reflection from the samples was compared to the response from the standard. Absolute measurements cannot be made by this method, of course, since the reflectivity is a function of particle size, packing of the sample, angles of illumination and viewing. Comparison of the results obtained with this system with true absorption measurements on materials obtainable in the form of single plates or crystals has, however, given good agreement as to the location of absorption bands and qualitative information as to their intensity. The difference in reflection between different samples in regions where they are both quite transparent is not to be considered significant on account of the disturbing factors mentioned above; significance should be attached only to the shape of each curve and the structure or absence of structure found therein. The measurements were made point by point, with data being taken every 50A between 2000 and 3400A. The data indicate that the introduction of manganese into calcite does not introduce any pronounced absorption beyond that present in pure calcite in the wave-length range investigated except at wave-lengths shorter than about 2400A. The introduction of lead in the calcite calls forth two fairly sharp absorption bands, which persist when manganese is introduced in addition to the lead.††† The absorption band lying at the longer wave-lengths in $CaCO_3$:Pb and $CaCO_3$(Pb+Mn) coincides fairly well with the position of the excitation band of the latter phosphor. The same situation applies in the case of $CaCO_3$:Tl and $CaCO_3$:(Tl+Mn). The results with cerium are less satisfactory because there is no sharp peak in

---

[16] C. F. Goodeve, Trans. Faraday Soc. 34, 902 (1938).

††† The greater absorption and overlapping of the peaks in $CaCO_3$:(Pb+Mn) compared to $CaCO_3$:Pb is due to the higher Pb concentration in the former, 1.60 mole percent *vs.* 0.4 mole percent Pb in the latter.

the absorption coinciding with the peak of the excitation spectrum of $CaCO_3$:(Ce+Mn). However, even here a considerable increase in absorption over pure $CaCO_3$ is evident, there being a drop of approximately 20–30 percent in the reflection between 3400A and 3150A for both $CaCO_3$:Ce and $CaCO_3$:(Ce+Mn).

A study of the literature reveals a number of known phosphors which appear to be based upon the mechanism of sensitized luminescence outlined above. Among these are alkaline earth borates, phosphates, silicates, and silico-phosphates,[17,18] and the NaCl:(Pb+Mn) of Murata and Smith.[7] In principle it should be possible to introduce excitation bands for any desired ultraviolet spectral region by choice of the proper sensitizer for any phosphor showing response to cathode-ray or x-ray excitation but not to ultraviolet excitation. Investigations under way at the Naval Research Laboratory are concerned with further studies along these lines.

## SUMMARY

It is shown that synthetic preparations of calcite, with manganese as an activator, are luminescent under cathode-ray excitation, but are substantially non-luminescent under ultraviolet irradiation. By co-precipitation of small amounts of lead, thallium, or cerium with the manganese-activated calcium carbonate, there is produced no essential change in the cathode-ray luminescence spectrum, but the products are then strongly excited by ultraviolet radiation as well. Spectrographic analyses of natural calcites from a number of different localities show that in the strongly luminescent (2537A excitation) specimens of the natural mineral, lead accompanies the manganese activator, while in the non-luminescent (2537A excitation) specimens the manganese is present without lead. The near-ultraviolet excited luminescence of most natural calcites is not connected with the lead, and has not as yet been ascribed to the presence of any specific auxiliary impurity.

The interpretation of these results is shown to be most consistent with the concept of a "sensitized" luminescence, and the possible application of this viewpoint to known phosphors and to the synthesis of new phosphors having specific excitation bands is pointed out.

## ACKNOWLEDGMENT

We wish to express our thanks to Mrs. Esther Claffy for the chemical analyses of our synthetic calcites. The chemical analysis of the Langban calcite was obtained through the kindness of Professor C. Palache, Department of Mineralogy, Harvard University.

---

[17] Hüniger and Panke, U. S. Patents 2,241,950 and 2,241,951 of May 13, 1941.

[18] Aschermann and Strubing, U. S. Patent 2,308,736 of January 19, 1943.

# Electron Beam Deflection

## Part I. Small-Angle Deflection Theory

R. G. E. HUTTER
*Sylvania Electric Products, Inc., Flushing, New York*

(Received February 26, 1947)

Two general mathematical methods are discussed which may be used to study the effects of electric and/or magnetic deflection fields on electron beams. The basic equations for the "path method" are the equations of motion of Newton or Euler-Lagrange. The "iconal method" makes use of certain properties of the Hamiltonian function.

These methods are then applied to describe the action of balanced, two-dimensional electric deflection fields on electron beams. It is shown that both methods yield essentially the same results. Expressions are derived describing the magnitude of deflection and the distortions of an electron beam.

Only the path method is used in a similar investigation of magnetic-type deflection fields.

## A. INTRODUCTION AND GENERAL OUTLINE OF THE THEORY

### 1. Introduction

THE problems of the formation and the control of electron beams are the concern of "Geometrical Electron Optics," a term chosen to imply the analogy which exists between the behavior of electron beams and light beams, and which manifests itself in the terminology and the fact that many light-optical instruments have an electron-optical equivalent. Electron beams may be focused, deflected, and reflected by electric or magnetic fields which are called electron lenses, deflection fields, and electron mirrors, respectively. The deflection fields might be termed electron prisms if it were not for the fact that deflection fields produce both prismatic *and* focusing effects on electron beams. A deflection field may be considered as a prism having a weak cylinder lens joined to each face.

If the deflection fields are classified as those making use of focusing effects and those for which these effects are undesirable, it is found that these two classes represent the deflection fields producing large-angle deflections and those producing small-angle deflections, respectively. Deflection fields of the former group are employed in mass spectrometers, cyclotrons, and microanalyzers, while fields of the latter group are used in cathode-ray and television tubes, ion traps, scanning-type electron microscopes, etc.

This paper will deal exclusively with small-angle deflection fields in the manner now used for the treatment of electron-optical phenomena; the terms "deflection field," and "deflection system" will hereafter be assumed to relate only to small-angle deflection fields or systems. Large-angle deflection fields have been discussed elsewhere [1-4]

It is the writer's purpose to provide certain mathematical tools which should be found useful in the design of improved deflection systems. In spite of the practical importance of an understanding of the problems of deflection, little attention has been paid to a systematic study of deflection fields. Approximate formulae are still used to determine the magnitude of deflection, and qualitative reasoning is applied to obtain information about the type of distortion of the deflected beam. Two exceptions exist: J. Picht and J. Himpan[5] have investigated single and crossed two-dimensional electric fields, and G. Wendt[6] has studied single and crossed magnetic deflection fields.

A thorough discussion of the subject is indicated for a number of reasons:

(1) Because of war conditions, the work of Picht, Himpan, and Wendt appears to have escaped the attention of the workers in the field of electron optics.

---

[1] R. G. E. Hutter, Phys. Rev. **67**, 248 (1945).

[2] R. Herzog, Zeits. f. Physik **89**, 447 (1934).

[3] W. Henneberg, Ann. d. Physik **19**, 335 (1934), *ibid.* **20**, 1 (1934), *ibid.* **21**, 390 (1934).

[4] N. D. Coggeshall, Phys. Rev. **70**, 270 (1946).

[5] J. Picht and J. Himpan, Ann. d. Physik **39**, 409 (1941), *ibid.* **43**, 53 (1943).

[6] G. Wendt, Die Telefunkenröhre **15**, 100 (1939); Zeits. f. Physik **118**, 593 (1942).

(2) These authors used a method of investigation which is known in electron optics as the "path method." It is of interest to show that another method, called the "iconal method," may also be employed.

(3) The method of Picht, Himpan, and Wendt, as well as the results, may be simplified. This simplification results in greater ease of application.

(4) The investigations may be generalized to include three-dimensional electric deflection fields.*

(5) A number of applications of the simplified and extended theory have now been made resulting in design charts for the determination of the magnitude of deflection, and in quantitative information on spot and pattern distortions of some of the commonly employed deflection fields.

The work is to be divided into two parts.** In Part I the general framework of the theory will be discussed, and the theory will be specialized for the case of single two-dimensional electric deflection fields, and single and crossed magnetic deflection fields. In Part II the magnitude of deflection and the distortion effects are computed for a number of typical electric and magnetic deflection fields.

## 2. General Outline of the Theory

Cathode-ray and television tubes are the most important devices employing the kinds of deflection fields to be investigated here. Although the theory to be developed may also be applied to small-angle deflection fields used in other devices, the terminology is that pertaining to cathode-ray tubes. The tube for which the theory will be developed is, however, an ideal rather than an actual cathode-ray tube. In order not to complicate the theory unduly, a number of assumptions had to be made chiefly concerning the electron beam and the focusing system.

The electron beam is assumed to be a rectilinear flow of electrons, homogeneous in velocity, and either conically or cylindrically shaped; space-charge effects are ignored. Such assumptions are not essential to the theory but were made to simplify its application. If it is necessary to explain the effects in an actual cathode-ray tube, such effects might have to be taken into account separately.

Furthermore, field distortions of the deflection fields due to the presence of other electrodes were not taken into account. This means that the deflection fields were taken as being determined solely by the current conductors and electrodes intended for the formation of the field.

These idealizations and approximations are permissible if one is interested in a comparative study of deflection fields, but not in an explanation of the performance of a particular cathode-ray tube which employs deflection fields.

The Cartesian system of coordinates will be used throughout this work. The $z$ axis is taken as the optical axis of the focusing system. If undeflected, the electrons of the beam will travel in the $+z$ direction. The $x$ direction is also called the "horizontal" direction while the $y$ direction is called the "vertical" direction.

The initial coordinates of an electron, whose path through the deflection field is to be deter-

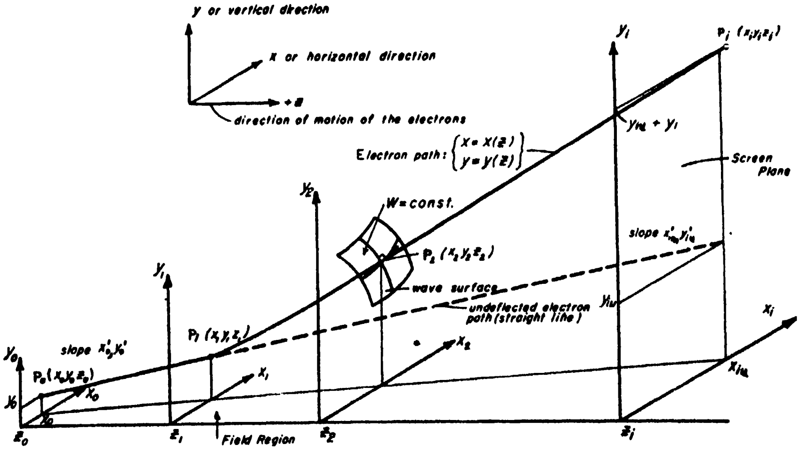


FIG. 1. The geometry of deflection.

---

* To be published at a later date.

** A part of the material presented in these papers was given in an abbreviated form at the National Electronics Conference in Chicago, October 3–5, 1946.

mined, are $z_0$, $x_0$, and $y_0$. The slope of the path at this point is given by $x_0'$, $y_0'$, the prime indicating the derivative of the quantity with respect to $z$. The position coordinates of an electron after deflection in a plane perpendicular to the $z$ axis at a point $z_t$ are given by $z_t$, $x_t$, and $y_t$. The coordinates $x_t$, $y_t$ at a specified coordinate $z_t$ are the quantities which ordinarily have to be determined. In cathode-ray and television tubes, $z_t$ is the axial coordinate of the plane of the fluorescent screen. For simplicity, it is assumed that the screen is flat, although in many tubes the screen is actually a curved surface. Since $z_t$ will remain a variable which may be chosen arbitrarily, it will also be possible to determine the position of the electron on a curved screen, in which case the calculations might be a little more cumbersome.

In most practical cases $z_0$ and $z_t$ are axial coordinates which are in regions where the influence of the deflection field is negligibly small. The deflection action of the field takes place between two $xy$ planes with axial coordinates $z_1$ and $z_2$, where $z_0<z_1<z_2<z_t$ (see Fig. 1).

There are four quantities—two coordinates and two slopes—which are of importance. If the deflecting field is zero, the electron with the above mentioned initial conditions $x_0$, $y_0$, $x_0'$, $y_0'$ at $z_0$, would proceed in a straight line from $x_0$, $y_0$, $z_0$ to the screen. Its coordinates at the screen are given by $x_{tu}$ and $y_{tu}$, where

$$x_{tu}=x_0+x_0'(z_t-z_0), \quad y_{tu}=y_0+y_0'(z_t-z_0). \tag{1}$$

Its slopes at that point are

$$x_{tu}'=x_0', \quad y_{tu}'=y_0'. \tag{2}$$

The second subscript $u$ indicates that $x_{tu}$, $y_{tu}$ are the coordinates of the undeflected electron at the screen. (See Fig. 1.)

Equations (1) and (2) may be used to express $x_0$, $y_0$ in terms of $x_{tu}$, $y_{tu}$, $x_{tu}'$, $y_{tu}'$ as follows:

$$x_0=x_{tu}-x_{tu}'(z_t-z_0), \quad y_0=y_{tu}-y_{tu}'(z_t-z_0). \tag{3}$$

The coordinates $x_t$, $y_t$ of the deflected electron at the axial coordinate $z_t$ may be expanded formally in power series of powers of $x_{tu}$, $y_{tu}$, and $x_{tu}'$, $y_{tu}'$:

$$\left.\begin{aligned} x_t &= \sum_{a,b,c,d} M_{a,b,c,d}\cdot x_{tu}^{a}\cdot y_{tu}^{b}\cdot x_{tu}'^{c}\cdot y_{tu}'^{d}, \\ y_t &= \sum_{a,b,c,d} N_{a,b,c,d}\cdot x_{tu}^{a}\cdot y_{tu}^{b}\cdot x_{tu}'^{c}\cdot y_{tu}'^{d}, \end{aligned}\right\} \tag{4}$$

where $a+b+c+d=n$ with $n=0, 1, 2, \cdots$. There are $[(n+1)(n+2)(n+3)]/6$ terms of degree $n$.

The coefficients $M_{a,b,c,d}$ and $N_{a,b,c,d}$ will, of course, be functions of $z_0$, $z_t$ and the field-strength parameters of the electric and magnetic deflection fields. The problem of determining $x_t$, $y_t$ reduces to one of determining these coefficients. It will be shown that the series, Eq. (4), are rapidly converging if small-angle deflections are considered only. In most practical cases, $n$ need not be taken as higher than three. The condition of smallness of the deflection angle will have to be stated mathematically.

Before going to the discussion of the two methods by which the coefficients $M_{a,b,c,d}$ and $N_{a,b,c,d}$ may be determined, a few remarks will be made concerning some general properties of these coefficients.

All electric and magnetic fields used as deflection fields are, strictly speaking, three-dimensional fields; i.e., the electrostatic potential function $\varphi$ and the magnetic field strength vector $\mathbf{H}$ will be functions of all three coordinates $x$, $y$, and $z$. However, certain fields employed in actual devices exhibit in certain regions the properties of two-dimensional fields. This is especially the case for some electric deflection fields. In this case the potential function $\varphi$ may be considered as depending only on $x$, $z$, or $y$, $z$. If the variation of $\varphi$ with the third coordinate is negligibly small in regions passed by the electron beam, a considerable simplification of the theory appears possible. This simplification finds its expression in the end result, Eq. (4), in which a number of the coefficients $M_{a,b,c,d}$ and $N_{a,b,c,d}$ will vanish. It is, however, unnecessary to prove which of the coefficients are zero, since the methods by which the coefficients are to be determined will automatically assign the value of zero to the proper coefficients.

Often, in practice, two deflection fields are used which are rotated 90 degrees about the $z$ axis with respect to each other. In this case only one set of coefficients $M_{a,b,c,d}$ or $N_{a,b,c,d}$ need be determined, the other set being obtained by interchanging $a$ and $b$ as well as $c$ and $d$, and interchanging the field quantities. In order to get $y_t$ from $x_t$, $x_{tu}$ must be replaced by $y_{tu}$, $x_{tu}'$ by $y_{tu}'$, and *vice versa*. It should be pointed out, however, that the numerical values of the corresponding coefficients will be different if different field functions are used for the two deflection fields.

The study of the motion of electrons in deflection fields requires a knowledge of the field distributions. An electric field may be given in terms of a potential distribution $\varphi=\varphi(x, y, z)$, while the magnetic field may be described by the magnetic vector potential distribution $\mathbf{A}=\mathbf{A}(x, y, z)$. It will be assumed that those functions are known.

Two essentially different methods are available to determine the position coordinates of an electron moving in the fields. (Both methods have been used to solve the analogous problem for focusing fields.)

If we consider electrons as small charged particles moving through electric and magnetic force fields, the concepts of electron ballistics can be applied. The Newton-Lorentz equations may be used as the basic set of equations. This method has been used by J. Picht and J. Himpan[5] for two-dimensional electrostatic deflection fields, and by G. Wendt[6] for a single magnetic deflection field. It is called the "path method."

If, on the other hand, waves are associated with a moving electron which travels through electric or magnetic fields of slowly-varying electron optical index of refraction, the wave approach to geometric optics can be applied. One of the basic concepts is that of the "point iconal." The method is based on the properties of the iconal function and is, therefore, called the "iconal method."

The common root of both methods is Hamilton's principle which states that the electron path,

$$x=x(s), \quad y=y(s), \quad z=z(s), \tag{5}$$

is such that

$$W=\int_{P_0(x_0,\,y_0,\,z_0)}^{P_i(x_i,\,y_i,\,z_i)} \left\{\left[\frac{2e}{m}\varphi(x, y, z)\right]^{\frac{1}{2}} - \frac{e}{m}(\mathbf{A}\cdot\mathbf{s})\right\} ds = \text{minimum}, \tag{6}$$

or with

$$\mathbf{s}=\frac{x'}{[1+x'^2+y'^2]^{\frac{1}{2}}}\mathbf{i}+\frac{y'}{[1+x'^2+y'^2]^{\frac{1}{2}}}\mathbf{j} +\frac{1}{[1+x'^2+y'^2]^{\frac{1}{2}}}\mathbf{k}, \tag{7}$$

and

$$ds=(1+x'^2+y'^2)^{\frac{1}{2}}dz, \tag{8}$$

that the electron path $x=x(z), y=y(z)$ is such that

$$W=\int_{z_0}^{z_i} F(x, y, z, x', y')dz =\int_{z_0}^{z_i} \left\{\left[\frac{2e}{m}\varphi(x, y, z)\right]^{\frac{1}{2}}\cdot[1+x'^2+y'^2]^{\frac{1}{2}} - \frac{e}{m}(A_x x'+A_y y'+A_z)\right\} dz=\text{minimum}, \tag{9}$$

where the primes indicate differentiation with respect to $z$.

The necessary conditions for the electron path Eq. (5) to satisfy Eq. (9) are certain differential equations. These equations are the Euler-Lagrange differential equations for the electron path obtained by carrying out the operations indicated by

$$\left.\begin{aligned}(d/dz)(\partial F/\partial x')-\partial F/\partial x=0,\\ (d/dz)(\partial F/\partial y')-\partial F/\partial y=0.\end{aligned}\right\} \tag{10}$$

These equations are equivalent to Newton's equations

$$m\frac{d^2\mathbf{r}}{dt^2}=e\mathbf{E}+e[\mathbf{v}\times\mathbf{H}], \tag{11}$$

with

$$\mathbf{r}\cdot\mathbf{i}+y\cdot\mathbf{j}+z\cdot\mathbf{k}. \tag{12}$$

The iconal method utilizes the properties of the function $W$. If $x(z)$ and $y(z)$ are taken as the functions describing the actual electron path, and if the initial point $P_0(x_0, y_0, z_0)$ is kept fixed, $W$ becomes a function of the end point $P_i(x_i, y_i, z_i)$. The equation $W=\text{const.}$ describes a surface about $P_0$. This is the wave surface of the electron radiation diverging from $P_0$. Assuming that we know the wave surface configuration in the space where electric and magnetic fields exist, the path of an electron beyond this space (with the boundary point $P_2(x_2, y_2, z_2)$) and into field-free regions to the point $P_i(x_i, y_i, z_i)$ is then given by[7] (see Fig. 1)

$$\left.\begin{aligned}x_i=x_2+(1/n_i)(\partial W/\partial x_2)[(x_i-x_2)^2\\ +(y_i-y_2)^2+(z_i-z_2)^2]^{\frac{1}{2}},\\ y_i=y_2+(1/n_i)(\partial W/\partial y_2)[(x_i-x_2)^2\\ +(y_i-y_2)^2+(z_i-z_2)^2]^{\frac{1}{2}}.\end{aligned}\right\} \tag{13}$$

---

[7] V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, and A. W. Vance, *Electron Optics and the Electron Microscope* (John Wiley and Sons, New York, 1945).

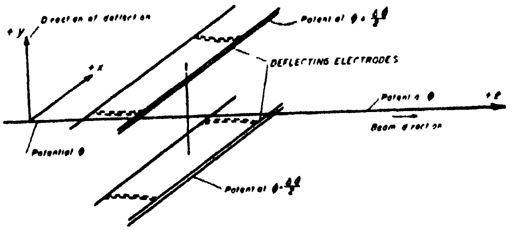


FIG. 2. The electrode geometry of a two-dimensional electric deflection field.

The slope of this straight line is that of the normal **n** to the wave surface at the point $P_2$, since **n** is given by

$$\mathbf{n}=\frac{1}{n_i}\frac{\partial W}{\partial x_2}\mathbf{i}+\frac{1}{n_i}\frac{\partial W}{\partial y_2}\mathbf{j}+\frac{1}{n_i}\frac{\partial W}{\partial z_2}\mathbf{k},$$
$$n_i=\phi_i^{\frac{1}{2}}=\left[(\partial W/\partial x_2)^2+(\partial W/\partial y_2)^2+(\partial W/\partial z_2)^2\right]^{\frac{1}{2}}. \tag{14}$$

It would appear that we have to know the electron path, i.e., the solution of the problem, in order to use the iconal method, since $W$ can only be determined if $x=x(z)$ and $y=y(z)$ are known. However, the restrictions on the magnitude of the deflection angle mentioned above, together with the property of $W$ as expressed by Eq. (9), eliminate the necessity of an exact knowledge of the path.

If the angle of deflection is small, the deflection itself must be small in the region where the field strengths are of appreciable magnitude. Large deflections are obtained by letting the beam travel in essentially field-free space before it reaches the screen; $x(z)$, $y(z)$, $x'(z)$, $y'(z)$ are, therefore, small quantities in regions of appreciable field strength. For their determination by means of Eq. (10), methods of successive approximations can be applied to yield electron paths of increasing accuracy. After the first-order approximation*** has been determined, either method may be used to find the higher order solutions.

In Part B, Section 2, the coordinates of the electron at the screen will be computed, including terms up to $n=2$ or $n=3$, which is, in general, sufficiently accurate to obtain quantitative information about the deflection, and the spot and pattern distortions occurring in common designs of cathode-ray and television tubes.

## B. TWO-DIMENSIONAL ELECTRIC DEFLECTION FIELDS

### 1. The Field Distribution

Electric fields which deflect the electron beam in cathode-ray and television tubes are ordinarily produced by metal electrodes connected to potential sources. The location of these electrodes with respect to the electron beam and the direction of deflection are shown in Fig. 2. The Cartesian system of coordinates is chosen so that $+z$ is in the direction of motion of the undeflected electrons, and $+y$ is in the direction of deflection. Two assumptions will be made about the shape and the potentials of the deflecting electrodes. These assumptions will be expressed in terms of the potential distribution $\varphi=\varphi(x, y, z)$ of the field produced by the deflecting system.

$$\varphi=\varphi(y, z)\quad \text{i.e., } \varphi \text{ is independent of } x, \tag{15}$$

$$\varphi(y, z)=\phi+\varphi^*(y, z),$$

where

$$-\varphi^*(+y, z)=\varphi^*(-y, z), \tag{16}$$

so that

$$\varphi(-y, z)=\phi-\varphi^*(+y, z)$$

where $\phi$ is the potential of the last anode of the electron gun.

The first assumption, Eq. (15), means that the electrodes producing $\varphi$ are considered to be of infinite extent in the $x$ direction, or, in other words, that the edge effects of the electrodes of finite extent in the $x$ direction are neglected. The error introduced by this assumption ordinarily is negligible, since most electrodes are of sufficient width in the $x$ direction. It follows from the assumption that the equipotential plots, as well as the cross section of the electrodes, are identical in all planes parallel to the $(y, z)$ plane.

The second assumption, Eq. (16), restricts the validity of the following investigations to so-called balanced fields.**** The electrodes are located symmetrically with respect to the center or $(x, z)$ plane. This center plane is at potential $\phi$. The potentials of the electrodes are $\phi+\frac{1}{2}\Delta\phi$ and $\phi-\frac{1}{2}\Delta\phi$, respectively; $\Delta\phi$ is the potential difference between the electrodes.

The potential function $\varphi$ must satisfy the

---

*** The first-order solution corresponds to the so-called "paraxial ray solution" of the focusing systems.

**** The theory of unbalanced deflection fields has also been worked out and will be published at a later date.

Laplace equation:

$$[\partial^2\varphi(y, z)/\partial y^2]+[\partial^2\varphi(y, z)/\partial z^2]\equiv 0. \tag{17}$$

Since $\varphi^*(y, z)=-\varphi^*(-y, z)$, the power series expansion of $\varphi^*$ in powers of $y$ can have only terms of odd power in $y$.

$$\begin{aligned}\varphi^*(y, z)&=a_1(z)\cdot y+a_3(z)\cdot y^3+\cdots\\&\quad+a_{2n+1}(z)\cdot y^{2n+1}+\cdots.\end{aligned} \tag{18}$$

Substituting Eq. (18) in Eq. (17), a recurrence formula for the coefficients is obtained, in which $a_1(z)$ remains arbitrary. With

$$a_1(z)=-E(z) \tag{19}$$

the following series is derived:

$$\begin{aligned}\varphi(y, z)&=\phi-E(z)\cdot y+\tfrac{1}{6}E''(z)y^3\\&\quad-(1/120)E^{(\mathrm{IV})}(z)y^5+\cdots(-1)^{n+1}\\&\quad\times[1/(2n+1)!]E^{(2n)}(z)\cdot y^{2n+1}+\cdots.\end{aligned} \tag{20}$$

It is seen that

$$E(z)=-\left.\frac{\partial\varphi(y, z)}{\partial y}\right|_{y=0}, \tag{21}$$

i.e., $E(z)$ is the axial field-strength distribution.

The electron-optical theory described below will be based on the general form of the potential distribution Eq. (20). It should be noted that edge effects of electrodes in the $z$ direction are not excluded, since $E(z)$ is left arbitrary. The axial field-strength distribution $E(z)$ determines, except for a constant, the equipotential distribution of deflection fields satisfying the condition expressed by Eq. (15).

Ordinarily, two fields of the type described by Eq. (20) are employed in cathode-ray and television tubes. The electrodes are rotated 90 degrees with respect to each other about the $z$ axis; one field deflects in the $y$ direction, the other in the $x$ direction. The two pairs of electrodes are generally spaced sufficiently far apart along the $z$ axis to prevent the distortion of the potential distribution of one field by the electrodes producing the other field. It is assumed here that a plane exists midway between the two fields where neither field has sufficient strength to affect the electrons. A study of crossed electric fields reduces, then, to a study of two single fields.

### 2. Path Method for a Single Two-Dimensional Electric Deflection Field

Integration of the path differential Eqs. (10) becomes possible if it is assumed that $x(z)$, $y(z)$, $x'(z)$, and $y'(z)$ are small quantities. The expressions $\varphi^{\frac{1}{2}}$ and $(1+x'^2+y'^2)^{\frac{1}{2}}$ of Eq. (9) may be written in series form

$$\varphi^{\frac{1}{2}}(y, z)=\phi^{\frac{1}{2}}\left(1-\frac{1}{2}\frac{E}{\phi}y+\frac{1}{12}\frac{E''}{\phi}y^3-\frac{1}{8}\frac{E^2}{\phi^2}y^2-\frac{1}{16}\frac{E^3}{\phi^3}y^3+\cdots\right), \tag{22}$$

and

$$(1+x'^2+y'^2)^{\frac{1}{2}}=1+\tfrac{1}{2}(x'^2+y'^2)-\tfrac{1}{8}(x'^2+y'^2)^2+\cdots. \tag{23}$$

The expression for $W$ then becomes

$$\left.\begin{aligned}W=\left(\frac{2e}{m}\phi\right)^{\frac{1}{2}}\int_{z_0}^{z_i}\Bigg\{&[1]-\left[\frac{1}{2}\frac{E}{\phi}y\right]+\left[\tfrac{1}{2}(x'^2+y'^2)-\frac{1}{8}\frac{E^2}{\phi^2}y^2\right]+\Bigg[-\tfrac{1}{4}(x'^2+y'^2)\frac{E}{\phi}y\\&+\left(\frac{1}{12}\frac{E''}{\phi}-\frac{1}{16}\frac{E^3}{\phi^3}\right)y^3\Bigg]+\Bigg[-\tfrac{1}{8}(x'^2+y'^2)^2-\tfrac{1}{16}(x'^2+y'^2)y^2\frac{E^2}{\phi^2}\\&-\left(-\frac{1}{24}\frac{EE''}{\phi^2}+\frac{5}{128}\frac{E^4}{\phi^4}\right)y^4\Bigg]+\text{higher order terms}\Bigg\}dz,\\&\text{or}\\W=\left(\frac{2e}{m}\phi\right)^{\frac{1}{2}}\int_{z_0}^{z_i}\{&F_0+F_1+F_2+F_3+F_4+\text{higher order terms}\}dz,\end{aligned}\right\} \tag{24}$$

if the bracketed expressions are called $F_0$, $F_1$, $F_2$, $F_3$, and $F_4$. The Euler-Lagrange differential equations

$$\left.\begin{aligned}(d/dz)[(\partial/\partial x')(F_1+F_2+F_3+F_4)]-0&=0,\\ (d/dz)[(\partial/\partial y')(F_1+F_2+F_3+F_4)]-(\partial/\partial y)(F_1+F_2+F_3+F_4)&=0,\end{aligned}\right\} \tag{25}$$

yield

$$\left.\begin{aligned}&\frac{d}{dz}\left[x'-\frac{1}{2}\frac{E}{\phi}yx'-\tfrac{1}{2}x'^3-\tfrac{1}{2}y'^2x'-\frac{1}{8}\frac{E^2}{\phi^2}y^2x'\right]=0,\\ &\frac{d}{dz}\left[y'-\frac{1}{2}\frac{E}{\phi}yy'-\tfrac{1}{2}y'^3-\tfrac{1}{2}x'^2y'-\frac{1}{8}\frac{E^2}{\phi^2}y^2y'\right]-\left[-\frac{1}{2}\frac{E}{\phi}-\frac{1}{4}\frac{E^2}{\phi^2}y-\frac{1}{4}\frac{E}{\phi}x'^2-\frac{1}{4}\frac{E}{\phi}y'^2\right.\\ &\qquad\left.-\frac{3}{16}\frac{E^3}{\phi^3}y^2+\frac{1}{4}\frac{E''}{\phi}y^2-\frac{1}{8}\frac{E^2}{\phi^2}x'^2y-\frac{1}{8}\frac{E^2}{\phi^2}y'^2y+\frac{1}{6}\frac{EE''}{\phi^2}y^3-\frac{5}{32}\frac{E^4}{\phi^4}y^3\right]=0.\end{aligned}\right\} \tag{26}$$

The restriction that the deflection be small where the field strength is of appreciable magnitude may be expressed mathematically as

$$\left|\frac{1}{2}\frac{E}{\phi}y\right|\ll 1. \tag{27}$$

Taking Eq. (27) and $|x'^2|\ll 1$ and $|y'^2|\ll 1$ into account, the approximate solution of Eq. (26) may then be obtained in two steps. To the first order of approximation, the electron path will be given by

$$(d/dz)(x')=0,\quad (d/dz)(y')=-\tfrac{1}{2}(E/\phi). \tag{28}$$

The solution of Eq. (28) is

$$\left.\begin{aligned}x_g(z)&=x_0+x_0'(z-z_0),\\ y_g(z)&=y_0+y_0'(z-z_0)+Y(z),\end{aligned}\right\}\dagger \tag{29}$$

where $x_0$, $x_0'$, $y_0$, and $y_0'$ are the initial conditions of the electron path. Here

$$Y(z)=-(1/2\phi)\int_{z_0}^{z}d\zeta\int_{z_0}^{\zeta}E(u)du, \tag{30}$$

is the deflection caused by the electrostatic field. If $z=z_1$ is the position coordinate of the screen, the deflection, $d$, at the screen for an electron with the initial conditions $x_0=y_0=x_0'=y_0'=0$ is given by

$$d=Y_1=Y(z_1)=-(1/2\phi)\int_{z_0}^{z_1}d\zeta\int_{z_0}^{\zeta}E(u)du. \tag{31}$$

Since $E(z)$ is proportional to the potential difference $\Delta\phi$ between the electrodes, it can be seen from Eq. (31) that $d$ is a linear function of $\Delta\phi$. This is a desirable property of a deflection field, but an actual field has this property only to the extent that it is permissible to neglect the higher order terms in Eq. (26). A more accurate solution of Eq. (26) will be given which shows the extent to which the actual paths deviate from the first-order paths.

The quantities $x'$, $y'$, $|\frac{1}{2}(E/\phi)y|$, are still assumed to be small, but not negligibly so. It may be expected that the actual solution of Eq. (26) will differ but little from the first-order solution Eq. (29). Only a small error will therefore be made if $y$ and $x$ of Eq. (29) are substituted in all terms of Eq. (26) which were previously neglected. The terms $x_0$, $x_0'$, $y_0$, and $y_0'$ would then appear in the new solutions for $x=x(z)$ and $y=y(z)$. It is, however, more convenient to use four other quantities, namely, the coordinates and slopes of the undeflected electron at the

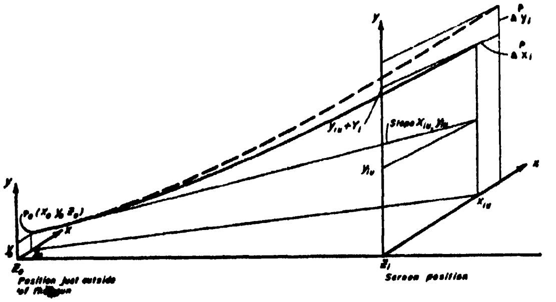


FIG. 3. The geometry of deflection (path method).

---

† This solution corresponds to the "paraxial ray" equation of the theory of focusing fields. In this theory the equation describes a path called the Gaussian path. It was for this reason that the subscript $g$ was chosen.

screen defined by Eq. (1). Replacing $x_0$, $y_0$, $x_0'$, and $y_0'$ in Eq. (29) by $x_{iu}$, $y_{iu}$, $x_{iu}'$, and $y_{iu}'$ (using Eqs. (2), (3)), the following equations are obtained for $x_g$ and $y_g$:

$$\left.\begin{aligned} x_g(z) &= x_{iu}+x_{iu}'(z-z_i), \\ y_g(z) &= y_{iu}+y_{iu}'(z-z_i)+Y(z). \end{aligned}\right\} \tag{32}$$

Substituting Eq. (32) in Eq. (26), the solutions $x(z)$ and $y(z)$ of the resulting differential equations are obtained by simple integrations. The terms of this solution may be arranged in powers of $x_{iu}$, $y_{iu}$, $x_{iu}'$, $y_{iu}'$. If the difference between the new and the former solution is taken and the integrals are extended from $z_0$ to $z_i$, two quantities $\Delta^P x_i$, $\Delta^P y_i$ are derived which describe the deviations at the screen of the actual electron path from the first-order path given by Eq. (29). (See Fig. 3.) These quantities $\Delta^P x_i$ and $\Delta^P y_i$ may be written as

$$\left.\begin{aligned} \Delta^P x_i &= \alpha_{0010}x_{iu}'+\alpha_{0110}x_{iu}'y_{iu}+\alpha_{0011}x_{iu}'y_{iu}'+\alpha_{0210}x_{iu}'y_{iu}^2+\alpha_{0111}x_{iu}'y_{iu}'y_{iu}+\alpha_{0012}x_{iu}'y_{iu}'^2, \\ \Delta^P y_i &= \beta_{0000}+\beta_{0100}y_{iu}+\beta_{0001}y_{iu}'+\beta_{0200}y_{iu}^2+\beta_{0020}x_{iu}'^2+\beta_{0002}y_{iu}'^2+\beta_{0101}y_{iu}'y_{iu} \\ &\quad +\beta_{0300}y_{iu}^3+\beta_{0120}x_{iu}'^2y_{iu}+\beta_{0102}y_{iu}y_{iu}'^2+\beta_{0021}x_{iu}'^2y_{iu}'+\beta_{0201}y_{iu}^2y_{iu}'+\beta_{0003}y_{iu}'^3. \end{aligned}\right\} \tag{33}$$

Here the notation, similar to that for the expressions $x_i$ and $y_i$, Eqs. (4), has been used to describe the expansion of the additional deflections $\Delta^P x_i$ and $\Delta^P y_i$. Since

$$x_i \cong x_g(z_i)+\Delta^P x_i = x_{iu}+\Delta^P x_i,$$

$$y_i \cong y_g(z_i)+\Delta^P y_i = y_{iu}+Y_i+\Delta^P y_i.$$

Then, to the degree of approximation considered here

$$M_{1000}=1+\alpha_{1000}, \quad \text{otherwise} \quad M_{a,b,c,d}\equiv\alpha_{a,b,c,d}$$

and

$$N_{0000}=\beta_{0000}+Y_i, \quad N_{0100}=1+\beta_{0100}, \quad \text{otherwise} \quad N_{a,b,c,d}\equiv\beta_{a,b,c,d}.$$

The coefficients $\alpha_{a,b,c,d}$ and $\beta_{a,b,c,d}$ are certain definite integrals from $z_0$ to $z_i$ which can be written in various forms. After carrying out the operations mentioned above, some of these coefficients contain double integrals which can be changed to single integrals by the use of the formula

$$\int_{z_0}^{z_i} d\zeta \int_{z_0}^{\zeta} F(u)du = -\int_{z_0}^{z_i} F(u)(u-z_i)du. \tag{34}$$

Some integrands have the second derivative of the axial field distribution $E''(z)$ as a factor. In general, $E(z)$ is known from experiment and $E''(z)$ can then only be obtained by numerical differentiation. This leads to inaccuracies which can be avoided by eliminating $E''(z)$ from all integrands employing the method of partial integration. This method and the equations

$$Y''(z) = -\frac{1}{2\phi}E(z), \tag{35}$$

and

$$Y_i = Y(z_i) = (1/2\phi)\int_{z_0}^{z_i}(\alpha-z_i)E(\alpha)d\alpha \tag{36}$$

were used to simplify the coefficients $\alpha_{a,b,c,d}$ and $\beta_{a,b,c,d}$ which were first derived by J. Picht and

J. Himpan[6] in a different form. They may be written as follows:

$$
\left.\begin{aligned}
\alpha_{0100} &= \tfrac{1}{2} Y_i Y_i' + \frac{3}{4\phi}\int_{z0}^{z_i} EY\,dz + \frac{1}{8\phi^2}\int_{z0}^{z_i} E^2 Y^2 dz, \\
\alpha_{0110} &= -\frac{1}{2\phi}E_0(z_i - z_0) - Y_i' + \frac{1}{4\phi^2}\int_{z0}^{z_i} E^2 Y dz, \\
\alpha_{0011} &= 2Y_i + \frac{1}{2\phi}E_0(z_i - z_0)^2 + \frac{1}{4\phi^2}\int_{z0}^{z_i} E^2 Y(z - z_i)dz, \\
\alpha_{0210} &= -\frac{1}{8\phi^2}E_0^2(z_i - z_0) + \frac{1}{8\phi^2}\int_{z0}^{z_i} E^2 dz, \\
\alpha_{0012} &= -\frac{1}{8\phi^2}E_0^2(z_i - z_0)^3 + \frac{1}{8\phi^2}\int_{z0}^{z_i} E^2 (z - z_i)^2 dz, \\
\alpha_{0111} &= \frac{1}{4\phi^2}E_0^2(z_i - z_0)^2 + \frac{1}{4\phi^2}\int_{z0}^{z_i} E^2 (z - z_i)dz;
\end{aligned}\right| \qquad (37)
$$

$$
\left.\begin{aligned}
\beta_{0000} = {} & \tfrac{1}{3} Y_i Y_i'^2 + \frac{1}{4\phi}E_i Y_i^2 - \frac{1}{6\phi}\int_{z0}^{z_i} EYY'dz + \frac{1}{2\phi^2}\int_{z0}^{z_i} E^2 Y(z - z_i)dz \\
& + \frac{1}{12\phi^2}E_i^2 Y_i^3 - \frac{3}{8\phi^2}\int_{z0}^{z_i} E^2 YY'^2(z - z_i)dz - \frac{3}{8\phi^2}\int_{z0}^{z_i} E^2 Y^2 Y'dz + \frac{1}{6\phi^2}\int_{z0}^{z_i} E'^2 Y^3(z - z_i)dz \\
& + \frac{5}{16\phi^3}\int_{z0}^{z_i} E^3 Y^2(z - z_i)dz + \frac{5}{32\phi^4}\int_{z0}^{z_i} E^1 Y^3(z - z_i)dz, \\
\beta_{0100} = {} & \tfrac{1}{2} Y_i'^2 + \frac{1}{2\phi}E_i Y_i + \frac{1}{4\phi^2}E_i^2 Y_i^2 + \frac{1}{2\phi^2}\int_{z0}^{z_i} E^2(z - z_i)dz - \frac{3}{8\phi^2}\int_{z0}^{z_i} E^2 Y'^2(z - z_i)dz \\
& + \frac{1}{2\phi^2}\int_{z0}^{z_i} E'^2 Y^2(z - z_i)dz - \frac{3}{4\phi^2}\int_{z0}^{z_i} E^2 YY'dz + \frac{5}{8\phi^3}\int_{z0}^{z_i} E^3 Y(z - z_i)dz \\
& + \frac{15}{32\phi^1}\int_{z0}^{z_i} E^1 Y^2(z - z_i)dz, \\
\beta_{0001} = {} & \tfrac{1}{2} Y_i Y_i' - \frac{1}{4\phi}\int_{z0}^{z_i} EY\,dz + \frac{1}{2\phi^2}\int_{z0}^{z_i} E^2(z - z_i)^2 dz - \frac{3}{8\phi^2}\int_{z0}^{z_i} E^2 Y^2 dz \\
& - \frac{3}{2\phi^2}\int_{z0}^{z_i} E^2 YY'(z - z_i)dz - \frac{3}{8\phi^2}\int_{z0}^{z_i} E^2 Y'^2(z - z_i)^2 dz + \frac{1}{2\phi^2}\int_{z0}^{z_i} E'^2 Y^2(z - z_i)^2 dz \\
& + \frac{5}{8\phi^3}\int_{z0}^{z_i} E^3 Y(z - z_i)^2 dz + \frac{15}{32\phi^4}\int_{z0}^{z_i} E^4 Y^2(z - z_i)^2 dz,
\end{aligned}\right|
$$

(*Equation 38 continued on following page*)

$$\left.\begin{aligned}
\beta_{0200}&=-\frac{1}{4\phi}[E_0'(z_i-z_0)+(E_0-E_i)]+\frac{1}{4\phi^2}E_i^2Y_i-\frac{3}{8\phi^2}\int_{z_0}^{z_i}E^2Y'dz+\frac{1}{2\phi^2}\int_{z_0}^{z_i}E'^2Y(z-z_i)dz\\
&\qquad+\frac{5}{16\phi^3}\int_{z_0}^{z_i}E^3(z-z_i)dz+\frac{15}{32\phi^4}\int_{z_0}^{z_i}E^4Y(z-z_i)dz,\\
\beta_{0020}&=Y_i+\frac{1}{8\phi^2}\int_{z_0}^{z_i}E^2Y(z-z_i)dz,\\
\beta_{0002}&=-\frac{1}{4\phi}E_0'(z_i-z_0)^3-\frac{1}{4\phi}E_0(z_i-z_0)^2-\frac{9}{8\phi^2}\int_{z_0}^{z_i}E^2Y(z-z_i)dz-\frac{9}{8\phi^2}\int_{z_0}^{z_i}E^2Y'(z-z_i)^2dz\\
&\qquad+\frac{1}{2\phi^2}\int_{z_0}^{z_i}E'^2Y(z-z_i)^3dz+\frac{5}{16\phi^3}\int_{z_0}^{z_i}E^3(z-z_i)^3dz+\frac{15}{32\phi^4}\int_{z_0}^{z_i}E^4Y(z-z_i)^3dz,\\
\beta_{0101}&=Y_i'+\frac{1}{2\phi}E_0'(z_i-z_0)^2+\frac{1}{2\phi}E_0(z_i-z_0)-\frac{3}{4\phi^2}\int_{z_0}^{z_i}E^2Ydz-\frac{3}{2\phi^2}\int_{z_0}^{z_i}E^2Y'(z-z_i)dz\\
&\qquad+\frac{1}{\phi^2}\int_{z_0}^{z_i}E'Y(z-z_i)^2dz+\frac{5}{8\phi^3}\int_{z_0}^{z_i}E^3(z-z_i)^2dz+\frac{15}{16\phi^4}\int_{z_0}^{z_i}E^4Y(z-z_i)^2dz,\\
\beta_{0300}&=-\frac{1}{6\phi^2}E_0E_0'(z_i-z_0)+\frac{1}{12\phi^2}(E_i^2-E_0^2)+\frac{1}{6\phi^2}\int_{z_0}^{z_i}E'^2(z-z_i)dz+\frac{5}{32\phi^4}\int_{z_0}^{z_i}E^4(z-z_i)dz,\\
\beta_{0120}&=\frac{1}{8\phi^2}\int_{z_0}^{z_i}E^2(z-z_i)dz,\\
\beta_{0102}&=-\frac{1}{2\phi^2}E_0^2(z_i-z_0)^2-\frac{1}{2\phi}E_0E_0'(z_i-z_0)^3-\frac{9}{8\phi^2}\int_{z_0}^{z_i}E^2(z-z_i)dz\\
&\qquad+\frac{1}{2\phi^2}\int_{z_0}^{z_i}E'^2(z-z_i)^3dz+\frac{15}{32\phi^4}\int_{z_0}^{z_i}E^4(z-z_i)^3dz,\\
\beta_{0021}&=\frac{1}{8\phi^2}\int_{z_0}^{z_i}E^2(z-z_i)^2dz,\\
\beta_{0201}&=\frac{3}{8\phi^2}E_0^2(z_i-z_0)+\frac{1}{2\phi^2}E_0E_0'(z_i-z_0)^2-\frac{3}{8\phi^2}\int_{z_0}^{z_i}E^2dz\\
&\qquad+\frac{1}{2\phi^2}\int_{z_0}^{z_i}E'^2(z-z_i)^2dz+\frac{15}{32\phi^4}\int_{z_0}^{z_i}E^4(z-z_i)^2dz,\\
\beta_{0003}&=\frac{5}{24\phi^2}E_0^2(z_i-z_0)^3+\frac{1}{6\phi^2}E_0E_0'(z_i-z_0)^4-\frac{3}{4\phi^2}\int_{z_i}^{z_i}E^2(z-z_i)^2dz\\
&\qquad+\frac{1}{6\phi^2}\int_{z_0}^{z_i}E'^2(z-z_i)^4dz+\frac{5}{32\phi^4}\int_{z_0}^{z_i}E^4(z-z_i)^4dz.
\end{aligned}\right\}\qquad(38)$$

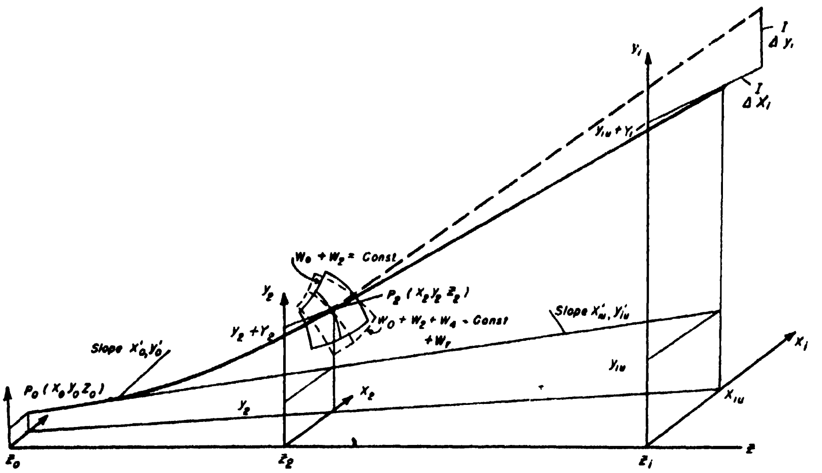


FIG. 4. The geometry of deflection (iconal method).

## 3. Iconal Method for a Single Electrostatic Deflection Field

Essentially the same coefficients can also be obtained by the use of the iconal method, as will now be shown. The point iconal function will be written in the form

$$W=\left(\phi\frac{2e}{m}\right)^{\frac{1}{2}}\int_{z_0}^{z_2}\left\{[1]+\left[-\frac{1}{2}\frac{E}{\phi}y+\tfrac{1}{2}(x'^2+y'^2)\right]+\left[-\frac{1}{8}\frac{E^2}{\phi^2}y^2-\tfrac{1}{8}(x'^2+y'^2)^2-\tfrac{1}{4}(x'^2+y'^2)y\frac{E}{\phi}\right.\right.$$
$$\left.+\frac{1}{12}\frac{E''}{\phi}y^3\right]+\left[-\frac{1}{16}\frac{E^3}{\phi^3}y^3-\tfrac{1}{16}(x'^2+y'^2)y^2\frac{E^2}{\phi^2}+\frac{1}{24}\frac{EE''}{\phi^2}y^4-\frac{5}{128}\frac{E^4}{\phi^4}y^4\right]$$
$$\left.+\text{higher order terms}\right\}dz, \tag{39}$$

or

$$W=W_0'+W_2+W_4+W_r+\text{higher order terms.}$$

If $P_0(x_0, y_0, z_0)$ remains fixed, $W=\text{const.}$ is the wave surface of electron radiation diverging from $P_0(x_0, y_0, z_0)$ and passing through the point $P_2(x_2, y_2, z_2)$. It is assumed that the space between the plane of the screen and the plane through $z_2$ parallel to the screen plane is field-free.†† (See Fig. 4.) The coordinates $x_i^{(1)}$ and $y_i^{(1)}$ of the electron path at the screen will then be given to the first degree approximation by

$$\left.\begin{aligned}x_i^{(1)}&=x_2+[(z_i-z_2)/\phi_i^{\frac{1}{2}}][\partial W_2/\partial x_2],\\ y_i^{(1)}&=y_2+[(z_i-z_2)/\phi_i^{\frac{1}{2}}][\partial W_2/\partial y_2].\end{aligned}\right\} \tag{40}$$

Equation (40) is derived from Eq. (13) by admitting only first-order terms. The expression $W_2$ has been determined by using the electron path equation in the form

$$\left.\begin{aligned}x_g(z)&=x_0[(z_2-z)/(z_2-z_0)]+x_2[(z-z_0)/(z_2-z_0)],\\ y_g(z)&=y_0[(z_2-z)/(z_2-z_0)]+y_2[(z-z_0)/(z_2-z_0)]-Y_2[(z-z_0)/(z_2-z_0)]+Y(z),\end{aligned}\right\} \tag{41}$$

with

$$Y(z)=-\frac{1}{2\phi}\int_{z_0}^{z}d\zeta\int_{z_0}^{\zeta}E(u)du. \tag{42}$$

The wave surface through the point $P_2(x_2, y_2, z_2)$, to a first approximation, is given by $W_0+W_2=\text{const.}$ More accurately, this surface will be given by $W_0+W_2+W_4+W_r=\text{const.}$ A new solution

---

†† This assumption can be discarded by an argument similar to that presented for focusing fields in the book by Zworykin, reference 7, p. 563.

for the coordinates of the electron path at the screen may be obtained in the form

$$\left.\begin{aligned} x_i^{(3)} &= x_2+\frac{z_i-z_2}{\phi_i^{\frac{1}{2}}}\frac{\partial W_2}{\partial x_2}\left[1+\frac{1}{2}\left(\frac{x_i^{(1)}-x_2}{z_i-z_2}\right)^2+\frac{1}{2}\left(\frac{y_i^{(1)}-y_2}{z_i-z_2}\right)^2\right]+\frac{z_i-z_2}{\phi_i^{\frac{1}{2}}}\frac{\partial(W_4+W_r)}{\partial x_2}, \\ y_i^{(3)} &= y_2+\frac{z_i-z_2}{\phi_i^{\frac{1}{2}}}\frac{\partial W_2}{\partial y_2}\left[1+\frac{1}{2}\left(\frac{x_i^{(1)}-x_2}{z_i-z_2}\right)^2+\frac{1}{2}\left(\frac{y_i^{(1)}-y_2}{z_i-z_2}\right)^2\right]+\frac{z_i-z_2}{\phi_i^{\frac{1}{2}}}\frac{\partial(W_4+W_r)}{\partial y_2}. \end{aligned}\right\} \quad (43)$$

The differences between the new and the former values of the coordinates are

$$\left.\begin{aligned} \Delta^I x_i &= x_i^{(3)}-x_i^{(1)}=\frac{z_i-z_2}{\phi_i^{\frac{1}{2}}}\frac{\partial(W_4+W_r)}{\partial x_2}+\frac{z_i-z_2}{\phi_i^{\frac{1}{2}}}\left[\frac{1}{2}\left(\frac{x_i^{(1)}-x_2}{z_i-z_2}\right)^2+\frac{1}{2}\left(\frac{y_i^{(1)}-y_2}{z_i-z_2}\right)^2\right]\frac{\partial W_2}{\partial x_2}, \\ \Delta^I y_i &= y_i^{(3)}-y_i^{(1)}=\frac{z_i-z_2}{\phi_i^{\frac{1}{2}}}\frac{\partial(W_4+W_r)}{\partial y_2}+\frac{z_i-z_2}{\phi_i^{\frac{1}{2}}}\left[\frac{1}{2}\left(\frac{x_i^{(1)}-x_2}{z_i-z_2}\right)^2+\frac{1}{2}\left(\frac{y_i^{(1)}-y_2}{z_i-z_2}\right)^2\right]\frac{\partial W_2}{\partial y_2}. \end{aligned}\right\} \quad (44)$$

Carrying out the indicated differentiations, the additional deflections $\Delta^I x_i$ and $\Delta^I y_i$ may be compared with the values $\Delta^P x_i$ and $\Delta^P y_i$ of the path method. The result of such a comparison is expressed by the following relations:

$$\left.\begin{aligned} \Delta^I x_i &= \Delta^P x_i-[(z_i-z_0)/(z_2-z_0)]\Delta^P x_2, \\ \Delta^I y_2 &= \Delta^P y_i-[(z_i-z_0)/(z_2-z_0)]\Delta^P y_2. \end{aligned}\right\} \quad (45)$$

Both methods give essentially the same result if the plane through $z_2$ terminates the deflection field where the deflection itself is still very small; then $\Delta^P x_2$ and $\Delta^P y_2$ will be small compared with $\Delta^P x_i$ and $\Delta^P y_i$, respectively.

The small difference between the results obtained by the two methods arises mainly from the fact that the path method corrects the first-order approximation path through the whole region from $z_0$ to $z_i$, while the iconal method corrects essentially only for the difference in slope at $z_2$, i.e., at the end of the field.

### 4. Some Remarks Concerning Applications

There are three problems which can be solved with the aid of the theory here presented. These are: (1) the determination of the "deflection" of the electron beam by any electric deflection system when the dimensions of the deflecting electrodes and their potentials are known; (2) the computation of the defocusing effects by a deflection system; and (3) the determination of deflection fields having less defocusing and distortion effects than those in common use.

The solution of the deflection problem requires a knowledge of the axial field-strength distribution $E(z)$. If the electrode shapes and potentials are known, $E(z)$ may, in some cases, be determined mathematically. In general, however, an experimental method must be used. The deflection may then be computed by means of Eq. (31). The problem of computing the defocusing and distortion effects requires a knowledge of $E(z)$, $E'(z)$, $Y(z)$, and $Y'(z)$. The defects of the deflection system are described by Eq. (33), the coefficients $\alpha_{a,b,c,d}$ and $\beta_{a,b,c,d}$ of Eqs. (37) and (38), ordinarily being determined by means of a planimeter. The coordinate $z_0$ is chosen to be outside the deflection field; $\alpha_{a,b,c,d}$ and $\beta_{a,b,c,d}$ are determined for any given screen position $z_i$ and any given potentials $\phi$ and $\Delta\phi$. Since $E(z)$ is proportional to $\Delta\phi$, the various terms in the coefficients may be arranged in powers of $\Delta\phi/\phi$. Terms other than the lowest in each coefficient may ordinarily be neglected, reducing the amount of work involved in making computations. In order to compute the spot distortion, $x_{iu}$, $y_{iu}$, and $x_{iu}'$, $y_{iu}'$ are chosen to be the coordinates and slopes, respectively, of an electron in the undeflected beam. It should be noted that if the electron beam is considered to be a rectilinear flow of electrons with sharp boundaries, either cylindrically or conically shaped, the computations need only be applied to a number of electrons on the circumference of the beam.

Equation (33) can also be used to determine pattern distortion. It has been pointed out that crossed electrostatic fields, necessary to obtain a

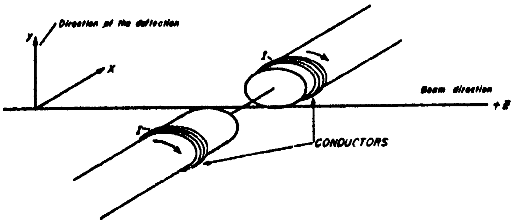


FIG. 5. Conductors and pole pieces producing a magnetic deflection field.

pattern, are arranged successively along the optical axis of the tube; it is assumed that a field-free plane exists between the two deflecting fields. The effects of both fields can then be superimposed linearly upon each other. An electron moving along the axis of the focusing system is deflected first by the vertical deflection system by the amount $Y(z_i)+\Delta^P y_i$ in the $y$ direction, and zero in the $x$ direction. Shortly after passing the vertical deflection field it passes through the horizontal deflection field, the effect of which is computed by replacing $x$ by $y$, and $y$ by $x$ in all equations written so far. The horizontal deflection will be

$$x_i=-\frac{1}{2\phi}\int_{z_0}^{z_i}d\zeta\int_{z_0}^{\zeta}E_h(u)du,$$

where $E_h(z)$ symbolizes the axial field-strength distribution of the horizontal deflection field. The quantities $\Delta^P y_i$ and $\Delta^P x_i$ determine the additional deflections when $x$ and $y$ are interchanged in Eq. (33). The "initial conditions" $x_{iu}$, $y_{iu}$, $x_{iu}'$, $y_{iu}'$ are given by

$$x_{iu}=0,\quad y_{iu}=Y_i,\quad x_{iu}'=0,\quad y_{iu}'=Y_i'.$$

Spot and pattern distortions were computed for a number of deflection systems and will be presented in Part II.

The third problem, namely, that of determining deflection fields having smaller defocusing and distortion effects than those ordinarily encountered, means, mathematically, to find a function $E(z)$ such that $Y_i$ has a prescribed value and yields coefficients $\alpha_{a,b,c,d}$ and $\beta_{a,b,c,d}$ smaller than those of known systems for the same value of $Y_i$. Various auxiliary conditions may be imposed such as those of constant electrode-to-screen distance, "reasonable" physical size of deflecting electrodes, etc. Solutions of these problems by mathematical procedures such as those used in the calculus of variations appear extremely difficult. It is more likely that trial and error methods, guided by the theory as expressed by Eqs. (31), (33), (37), and (38), will lead to improved deflection fields.

## C. MAGNETIC DEFLECTION FIELDS

### 1. The Field Distribution

Magnetic deflection fields of cathode-ray and television tubes are produced by a pair of iron or air-cored electromagnets placed in "ear-muff" position on the outside of the tube. This is illustrated in Fig. 5 by a single deflection field. In order to obtain a pattern on the screen of the tube, two such fields are needed, rotated 90 degrees about the optical axis with respect to each other. Ordinarily these two fields exist over the same interval along the axis. Electrostatic deflection fields were arranged one after the other along the axis. This difference between the arrangements necessitates a difference in theoretical treatment. The effects of the crossed magnetic fields can only be described by treating them as a single field obtained by superposition of the two fields, since both magnetic fields act simultaneously on the electron beam.

Series expressions for magnetic field distributions will now be derived in a manner similar to that used for electrostatic fields. Fields of the type pictured in Fig. 5 show a certain kind of symmetry. Again using Cartesian coordinates $x$, $y$, and $z$, with $+z$ in the direction of the original motion of the electrons, and with $y$ and $x$ in the direction of vertical and horizontal deflections, respectively, a mirror-type symmetry of the fields exists with respect to the $(x, z)$ and $(y, z)$ planes. The lines of force are mirror-symmetrical with respect to these planes. This condition may be expressed mathematically for the two fields by the following two sets of equations.

(1) Horizontal deflection field:

$$\left.\begin{aligned}H_x^h(x,y,z)&=-H_x^h(-x,y,z)\\&=-H_x^h(x,-y,z),\\H_y^h(x,y,z)&=+H_y^h(-x,y,z)\\&=+H_y^h(x,-y,z),\\H_z^h(x,y,z)&=+H_z^h(-x,y,z)\\&=-H_z^h(x,-y,z).\end{aligned}\right\}\quad(46)$$

(2) Vertical deflection field:

$$\left.\begin{aligned} H_x^v(x, y, z) &= +H_x^v(-x, y, z) \\ &= +H_x^v(x, -y, z), \\ H_y^v(x, y, z) &= -H_y^v(-x, y, z) \\ &= -H_y^v(x, -y, z), \\ H_z^v(x, y, z) &= -H_z^v(-x, y, z) \\ &= +H_z^v(x, -y, z), \end{aligned}\right\} \quad (47)$$

where $H_x^h$, $H_y^h$, $H_z^h$, and $H_x^v$, $H_y^v$, $H_z^v$ are the components of the magnetic field-strength vectors $\mathbf{H}^h$ and $\mathbf{H}^v$, respectively.

In current-free regions the relations

$$\operatorname{div}\mathbf{A}=0, \quad \nabla^2\mathbf{A}\equiv 0 \quad (48)$$

must hold for the magnetic vector potential $\mathbf{A}$, which is related to the magnetic field strength $\mathbf{H}$ by

$$\mathbf{H}=\operatorname{curl}\mathbf{A}. \quad (49)$$

The functions describing the components of the magnetic field strength may be expanded in power series of powers in $x$ and $y$ in which the coefficients are functions of $z$. Hence for the vertical deflection field one may write

$$\left.\begin{aligned} H_x^v &= p_{00}+p_{20}x^2+p_{02}y^2+p_{22}x^2y^2+p_{40}x^4+p_{04}y^4+p_{44}x^4y^4+p_{24}x^2y^4+p_{42}x^4y^2+\cdots, \\ H_y^v &= q_{11}xy+q_{13}xy^3+q_{31}x^3y+q_{33}x^3y^3+\cdots, \\ H_z^v &= r_{10}x+r_{12}xy^2+r_{32}x^3y^2+r_{14}xy^4+r_{34}x^3y^4+r_{30}x^3+\cdots. \end{aligned}\right\} \quad (50)$$

The procedure for the horizontal deflection field is the same.

In Eq. (50) use was made of the symmetry relations expressed by Eq. (47). The series for the components of the magnetic vector potential must then be (to give a relation expressed by (49)) of the form

$$\left.\begin{aligned} A_x^v &= a_{11}xy+a_{31}x^3y+a_{13}xy^3+a_{33}x^3y^3+a_{15}xy^5+\cdots, \\ A_y^v &= b_{00}+b_{20}x^2+b_{02}y^2+b_{22}x^2y^2+b_{40}x^4+b_{04}y^4+b_{24}x^2y^4+\cdots, \\ A_z^v &= c_{01}y+c_{21}x^2y+c_{03}y^3+c_{41}x^4y+c_{23}x^2y^3+c_{05}y^5+c_{43}x^4y^3+c_{25}x^2y^5+\cdots. \end{aligned}\right\} \quad (51)$$

The relations between the two sets of coefficients ($p_{\alpha\beta}$, $q_{\alpha\beta}$, $r_{\alpha\beta}$) and ($a_{\alpha\beta}$, $b_{\alpha\beta}$, $c_{\alpha\beta}$) may be found by the use of Eq. (49). Two more sets of relations between the coefficients of the set ($a_{\alpha\beta}$, $b_{\alpha\beta}$, $c_{\alpha\beta}$) may be found by use of the Eq. (48).

Redefining these coefficients $p_{00}$, $p_{02}$, and $p_{04}$ by

$$p_{00}\equiv -V_0(z), \quad p_{02}\equiv -V_2(z), \quad p_{04}\equiv -V_4(z), \quad (52)$$

the following series for the components $H_x^v$, $H_y^v$, and $H_z^v$ are obtained:

$$\left.\begin{aligned} H_x^v &= -V_0+(V_2+\tfrac{1}{2}V_0'')x^2-V_2y^2-[V_4+\tfrac{1}{6}V_2''+(1/24)V_0^{(\mathrm{IV})}]x^4 \\ &\qquad +(6V_4+\tfrac{1}{2}V_2'')x^2y^2-V_4y^4+\cdots, \\ H_y^v &= -2V_2xy-4V_4xy^3+(4V_4+\tfrac{1}{3}V_2'')x^3y+\cdots, \\ H_z^v &= -V_0'x-V_2'xy^2+(\tfrac{1}{3}V_2'+\tfrac{1}{6}V_0''')x^3+\cdots. \end{aligned}\right\} \quad (53)$$

The series for the components of the horizontal deflection field may be obtained by interchanging $x$ and $y$ in Eq. (53) and replacing $V_0$ by $-H_0$, $V_2$ by $-H_2$, and $V_4$ by $-H_4$. This yields

$$\left.\begin{aligned} H_x^h &= 2H_2xy+4H_4x^3y-(4H_4+\tfrac{1}{3}H_2'')xy^3+\cdots, \\ H_y^h &= H_0-(H_2+\tfrac{1}{2}H_0'')y^2+H_2x^2+[H_4+\tfrac{1}{6}H_2''+(1/24)H_0^{(\mathrm{IV})}]y^4 \\ &\qquad -(6H_4+\tfrac{1}{2}H_2'')x^2y^2+H_4x^4+\cdots, \\ H_z^h &= H_0'y+H_2'x^2y-(\tfrac{1}{3}H_2'+\tfrac{1}{6}H_0''')y^3+\cdots. \end{aligned}\right\} \quad (54)$$

The field strength of the combined fields is then given by

$$\left.\begin{aligned}
H_x &= -V_0+(V_2+\tfrac{1}{2}V_0'')x^2+2H_2xy-V_2y^2-[V_4+\tfrac{1}{6}V_2''+(1/24)V_0^{(\mathrm{IV})}]x^4+4H_4x^3y \\
&\qquad +(6V_4+\tfrac{1}{2}V_2'')x^2y^2-(4H_4+\tfrac{1}{3}H_2'')xy^3-V_4y^4+\cdots, \\
H_y &= H_0+H_2x^2-2V_2xy-(H_2+\tfrac{1}{2}H_0'')y^2+H_4x^4+(4V_4+\tfrac{1}{3}V_2'')x^3y \\
&\qquad -(6H_4+\tfrac{1}{2}H_2'')x^2y^2-4V_4xy^3+[H_4+\tfrac{1}{6}H_2''+(1/24)H_0^{(\mathrm{IV})}]y^4+\cdots, \\
H_z &= -V_0'x+H_0'y-V_2'xy^2+H_2'x^2y+(\tfrac{1}{3}V_2'+\tfrac{1}{6}V_0''')x^3-(\tfrac{1}{3}H_2'+\tfrac{1}{6}H_0''')y^3+\cdots.
\end{aligned}\right\} \tag{55}$$

The functions $-V_0$ and $+H_0$ are the axial field-strength distributions of the vertical and horizontal field strengths, respectively.

## 2. Theory of Magnetic Deflection

Either the path method or the iconal method may be used to describe the deflection and deflection defocusing effects of magnetic fields. Only the path method will be presented here; both methods give substantially the same result.

The equations of motion of an electron through any magnetic field are

$$\left.\begin{aligned}
\sqrt{\phi}(d/dz)[x'\ (1+x'^2+y'^2)^{\frac{1}{2}}] &= \eta[-y'H_z+H_y], \\
\sqrt{\phi}(d/dz)[y'/(1+x'^2+y'^2)^{\frac{1}{2}}] &= \eta[x'H_z-H_x],
\end{aligned}\right\} \tag{56}$$

where $\phi$ is the potential of the space in which the electrons are moving and $\eta=(e/2m)^{\frac{1}{2}}$, where $e$ and $m$ are charge and mass, respectively, of the electron. Equations (56) may be obtained either from

$$\delta\int_{P_0}^{P_1}\{[2(e/m)\phi]^{\frac{1}{2}}-e/m(\mathbf{A}\cdot\mathbf{s})\}ds=0, \tag{57}$$

with $\mathbf{H}=\mathrm{curl}\mathbf{A}$, or from Newton's equations of motion, written

$$m\frac{d^2\mathbf{r}}{dt^2}=-e[\mathbf{v}\times\mathbf{H}],\quad \text{with}\quad \mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k}. \tag{58}$$

If the magnetic field is that given by Eqs. (55), the Eqs. (56) become

$$\left.\begin{aligned}
(d/dz)[x'-\tfrac{1}{2}x'^3-\tfrac{1}{2}x'y'^2] &= (\eta/\sqrt{\phi})[V_0'y'x-H_0'y'y+H_0+H_2x^2-2V_2xy-(H_2+\tfrac{1}{2}H_0'')y^2], \\
(d/dz)[y'-\tfrac{1}{2}y'^3-\tfrac{1}{2}y'x'^2] &= (\eta/\sqrt{\phi})[-V_0'x'x+H_0'x'y+V_0-(V_2+\tfrac{1}{2}V_0'')x^2-2H_2xy+V_2y^2],
\end{aligned}\right\} \tag{59}$$

where only terms of zero, first and second order of $x$, $x'$, $y$, and $y'$ are considered. To a first approximation, the equations of motion are

$$x''=(\eta/\sqrt{\phi})H_0,\quad y''=(\eta/\sqrt{\phi})V_0. \tag{60}$$

The solution of these equations is

$$x_g(z)=x_0+x_0'(z-z_0)+X(z),\quad y_g(z)=y_0+y_0'(z-z_0)+Y(z), \tag{61}$$ †††

where

$$X(z)=\frac{\eta}{\sqrt{\phi}}\int_{z0}^{z}d\zeta\int_{z0}^{\zeta}H_0(u)du,\quad Y(z)=\frac{\eta}{\sqrt{\phi}}\int_{z0}^{z}d\zeta\int_{z0}^{\zeta}V_0(u)du. \tag{62}$$

The quantities $X=X(z)$ and $Y=Y(z)$ at $z=z_s$ (the screen position) are the deflections at the screen.

---

††† This solution corresponds to the paraxial ray equation of the theory of focusing fields.

The deflections in the two directions $x$ and $y$ are independent of each other in this first-order approximation.

It was assumed that the quadratic terms in Eq. (59) were negligibly small. They are still assumed to be small, but will now be taken into account. Following the procedure used for electric deflection fields, $x$, $x'$, $y$, and $y'$ of Eq. (59) will be replaced by $x_0$, $x_0'$, $y_0$, and $y_0'$ of Eq. (61). This yields

$$\left.\begin{aligned}
(d/dz)[x'-\tfrac{1}{2}x_0'^3-\tfrac{1}{2}x_0'y_0'^2]=(\eta/\sqrt{\phi})\{&V_0'[y_0'+Y'][x_0+x_0'(z-z_0)+X]\\
&-H_0'[y_0'+Y'][y_0+y_0'(z-z_0)+Y]+H_0+H_2[x_0^2+x_0'^2(z-z_0^2)+X^2+2x_0x_0'(z-z_0)\\
&+2x_0X+2x_0'(z-z_0)X]-2V_2[x_0y_0+x_0'y_0(z-z_0)+y_0X+x_0y_0'(z-z_0)\\
&+x_0'y_0'(z-z_0)^2+y_0'(z-z_0)X+x_0Y+x_0'(z-z_0)Y+XY]\\
&-(H_2+\tfrac{1}{2}H_0'')[y_0^2+y_0'^2(z-z_0)^2+Y^2+2y_0y_0'(z-z_0)+2y_0Y+2y_0'(z-z_0)Y]\},\\
(d/dz)[y'-\tfrac{1}{2}y_0'^3-\tfrac{1}{2}y_0'x_0'^2]=(\eta/\sqrt{\phi})\{&-V_0'[x_0'+X'][x_0+x_0'(z-z_0)+X]\\
&+H_0'[x_0'+X'][y_0+y_0'(z-z_0)+Y]+V_0-(V_2+\tfrac{1}{2}V_0'')[x_0^2+x_0'^2(z-z_0)^2+X^2\\
&+2x_0x_0'(z-z_0)+2x_0X+2x_0'(z-z_0)X]-2H_2[x_0y_0+x_0'y_0(z-z_0)+y_0X\\
&+x_0y_0'(z-z_0)+x_0'y_0'(z-z_0)^2+y_0'(z-z_0)X+x_0Y+x_0'(z-z_0)Y+XY]\\
&+V_2[y_0^2+y_0'^2(z-z_0)^2+Y^2+2y_0y_0'(z-z_0)+2y_0Y+2y_0'(z-z_0)Y]\}.
\end{aligned}\right\}\quad(63)$$

The initial conditions $x_0$, $x_0'$, $y_0$, and $y_0'$ may be replaced by $x_{iu}$, $x_{iu}'$, $y_{iu}$, and $y_{iu}'$ of the undeflected electron. The relation between these quantities is

$$x_0=x_{iu}+x_{iu}'(z_0-z_i),\quad y_0=y_{iu}+y_{iu}'(z_0-z_i),\quad x_0'=x_{iu}',\quad y_0'=y_{iu}'.\quad(64)$$

The solution of the differential equations, Eqs. (63) can be obtained by simple integrations. If the differences $\Delta^M x_i$ and $\Delta^M y_i$ between the new solution and that given by Eqs. (61) are taken, one obtains

$$\left.\begin{aligned}
\Delta^M x_i=\gamma_{0000}&+\gamma_{1000}x_{iu}+\gamma_{0100}y_{iu}+\gamma_{0010}x_{iu}'+\gamma_{0001}y_{iu}'+\gamma_{2000}x_{iu}^2\\
&+\gamma_{1100}x_{iu}y_{iu}+\gamma_{0200}y_{iu}^2+\gamma_{0020}x_{iu}'^2+\gamma_{0011}x_{iu}'y_{iu}'+\gamma_{0110}x_{iu}'y_{iu}\\
&+\gamma_{1001}x_{iu}y_{iu}'+\gamma_{1010}x_{iu}x_{iu}'+\gamma_{0101}y_{iu}y_{iu}'+\gamma_{0002}y_{iu}'^2,\\
\Delta^M y_i=\delta_{0000}&+\delta_{1000}x_{iu}+\delta_{0100}y_{iu}+\delta_{0010}x_{iu}'+\delta_{0001}y_{iu}'+\delta_{2000}x_{iu}^2\\
&+\delta_{1100}x_{iu}y_{iu}+\delta_{0200}y_{iu}^2+\delta_{0020}x_{iu}'^2+\delta_{0011}x_{iu}'y_{iu}'+\delta_{0110}x_{iu}'y_{iu}\\
&+\delta_{1001}x_{iu}y_{iu}'+\delta_{1010}x_{iu}x_{iu}'+\delta_{0101}y_{iu}y_{iu}'+\delta_{0002}y_{iu}'^2.
\end{aligned}\right\}\quad(65)$$

The coefficients $\gamma_{a,b,c,d}$ and $\delta_{a,b,c,d}$ in Eq. (65) are

$$\begin{aligned}
\gamma_{0000}=\tfrac{1}{2}\int X'^3dz+\tfrac{1}{2}\int X'Y'^2dz-(\eta/\sqrt{\phi})\int Y'XV_0'(z-z_i)dz&\\
+(\eta/\sqrt{\phi})\int YY'H_0(z-z_i)dz-(\eta/\sqrt{\phi})\int H_2X^2(z-z_i)dz&\\
+(\eta/\sqrt{\phi})\int Y^2H_2(z-z_i)dz+(2\eta/\sqrt{\phi})\int XYV_2(z-z_i)dz+\tfrac{1}{2}(\eta/\sqrt{\phi})\int Y^2H_0''(z-z_i)dz&,\\
\gamma_{1000}=-(2\eta/\sqrt{\phi})\int XH_2(z-z_i)dz+(2\eta/\sqrt{\phi})\int YV_2(z-z_i)dz-(\eta/\sqrt{\phi})\int Y'V_0'(z-z_i)dz&,
\end{aligned}$$

*(Equation 66 continued on following page)*

$$
\left.\begin{aligned}
\gamma_{0100}={}&(\eta/\sqrt{\phi})\int Y'H_0'(z-z_i)dz+(2\eta/\sqrt{\phi})\int XV_2(z-z_i)dz\\
&+(2\eta/\sqrt{\phi})\int YH_2(z-z_i)dz+(\eta/\sqrt{\phi})\int YH_0''(z-z_i)dz,\\
\gamma_{0010}={}&\tfrac{3}{2}\int X'^2dz+\tfrac{1}{2}\int Y'^2dz-(\eta/\sqrt{\phi})\int Y'V_0'(z-z_i)^2dz\\
&-(2\eta/\sqrt{\phi})\int XH_2(z-z_i)^2dz+(2\eta/\sqrt{\phi})\int YV_2(z-z_i)^2dz,\\
\gamma_{0001}={}&\int X'Y'dz-(\eta/\sqrt{\phi})\int XV_0'(z-z_i)dz\\
&+(\eta/\sqrt{\phi})\int YH_0'(z-z_i)dz+(\eta/\sqrt{\phi})\int Y'H_0'(z-z_i)^2dz+(2\eta/\sqrt{\phi})\int XV_2(z-z_i)^2dz\\
&+(2\eta/\sqrt{\phi})\int YH_2(z-z_i)^2dz+(\eta/\sqrt{\phi})\int YH_0''(z-z_i)^2dz,\\
\gamma_{2000}={}&-(\eta/\sqrt{\phi})\int H_2(z-z_i)dz,\\
\gamma_{1100}={}&(2\eta/\sqrt{\phi})\int V_2(z-z_i)dz,\\
\gamma_{0200}={}&(\eta/\sqrt{\phi})\int H_2(z-z_i)dz+\tfrac{1}{2}(\eta/\sqrt{\phi})\int H_0''(z-z_i)dz,\\
\gamma_{0020}={}&\tfrac{3}{2}\int X'dz-(\eta/\sqrt{\phi})\int H_2(z-z_i)^3dz,\\
\gamma_{0011}={}&\int Y'dz+(2\eta/\sqrt{\phi})\int V_2(z-z_i)^3dz-(\eta/\sqrt{\phi})\int V_0'(z-z_i)^2dz,\\
\gamma_{0110}={}&(2\eta/\sqrt{\phi})\int V_2(z-z_i)^2dz,\\
\gamma_{1001}={}&-(\eta/\sqrt{\phi})\int V_0'(z-z_i)dz+(2\eta/\sqrt{\phi})\int V_2(z-z_i)^2dz,\\
\gamma_{1010}={}&-(2\eta/\sqrt{\phi})\int H_2(z-z_i)^2dz,\\
\gamma_{0101}={}&(\eta/\sqrt{\phi})\int H_0'(z-z_i)dz+(2\eta/\sqrt{\phi})\int H_2(z-z_i)^2dz+(\eta/\sqrt{\phi})\int H_0''(z-z_i)^2dz,\\
\gamma_{0002}={}&\tfrac{1}{2}\int X'dz+(\eta/\sqrt{\phi})\int H_0'(z-z_i)^2dz+(\eta/\sqrt{\phi})\int H_2(z-z_i)^3dz+\tfrac{1}{2}(\eta/\sqrt{\phi})\int H_0''(z-z_i)^3dz;
\end{aligned}\right\}\quad(66)
$$

$$
\left.\begin{aligned}
\delta_{0000}={}&\tfrac{1}{2}\int Y'^3dz+\tfrac{1}{2}\int X'^2Y'dz\\
&+(\eta/\sqrt{\phi})\int XX'V_0'(z-z_i)dz-(\eta/\sqrt{\phi})\int X'YH_0'(z-z_i)dz\\
&+(\eta/\sqrt{\phi})\int X^2V_2(z-z_i)dz+\tfrac{1}{2}(\eta/\sqrt{\phi})\int X^2V_0''(z-z_i)dz\\
&\qquad+(2\eta/\sqrt{\phi})\int XYH_2(z-z_i)dz-(\eta/\sqrt{\phi})\int Y^2V_2(z-z_i)dz,\\
\delta_{1000}={}&(2\eta/\sqrt{\phi})\int XV_2(z-z_i)dz+(\eta/\sqrt{\phi})\int XV_0''(z-z_i)dz\\
&\qquad+(2\eta/\sqrt{\phi})\int YH_2(z-z_i)dz+(\eta/\sqrt{\phi})\int X'V_0'(z-z_i)dz,\\
\delta_{0100}={}&-(\eta/\sqrt{\phi})\int X'H_0'(z-z_i)dz+(2\eta/\sqrt{\phi})\int XH_2(z-z_i)dz-(2\eta/\sqrt{\phi})\int YV_2(z-z_i)dz,\\
\delta_{0010}={}&\int X'Y'dz+(\eta/\sqrt{\phi})\int XV_0'(z-z_i)dz+(\eta/\sqrt{\phi})\int X'V_0'(z-z_i)^2dz\\
&-(\eta/\sqrt{\phi})\int YH_0'(z-z_i)dz+(2\eta/\sqrt{\phi})\int XV_2(z-z_i)^2dz\\
&\qquad+(\eta/\sqrt{\phi})\int XV_0''(z-z_i)^2dz+(2\eta/\sqrt{\phi})\int YH_2(z-z_i)^2dz,\\
\delta_{0001}={}&\tfrac{3}{2}\int Y'^2dz+\tfrac{1}{2}\int X'^2dz-(\eta/\sqrt{\phi})\int X'H_0'(z-z_i)^2dz\\
&\qquad+(2\eta/\sqrt{\phi})\int XH_2(z-z_i)^2dz-(2\eta/\sqrt{\phi})\int YV_2(z-z_i)^2dz,\\
\delta_{2000}={}&(\eta/\sqrt{\phi})\int V_2(z-z_i)dz+\tfrac{1}{2}(\eta/\sqrt{\phi})\int V_0''(z-z_i)dz,\\
\delta_{1100}={}&(2\eta/\sqrt{\phi})\int H_2(z-z_i)dz,\\
\delta_{0200}={}&-(\eta/\sqrt{\phi})\int V_2(z-z_i)dz,\\
\delta_{0020}={}&\tfrac{1}{2}\int Y'dz+(\eta/\sqrt{\phi})\int V_0'(z-z_i)^2dz+(\eta/\sqrt{\phi})\int V_0(z-z_i)^3dz+\tfrac{1}{2}(\eta/\sqrt{\phi})\int V_0''(z-z_i)^3dz,
\end{aligned}\right| \tag{67}
$$

*(Equation 67 continued on following page)*

$$\delta_{0011}=\int X'dz-(\eta/\sqrt{\phi})\int H_0'(z-z_s)^2dz+(2\eta/\sqrt{\phi})\int H_2(z-z_s)^3dz,$$

$$\delta_{0110}=-(\eta/\sqrt{\phi})\int H_0'(z-z_s)dz+(2\eta/\sqrt{\phi})\int H_2(z-z_s)^2dz,$$

$$\delta_{1001}=(2\eta/\sqrt{\phi})\int H_2(z-z_s)^2dz,$$

$$\delta_{1010}=(\eta/\sqrt{\phi})\int V_0'(z-z_s)dz+(2\eta/\sqrt{\phi})\int V_2(z-z_s)^2dz+(\eta/\sqrt{\phi})\int V_0''(z-z_s)^2dz,$$

$$\delta_{0101}=-(2\eta/\sqrt{\phi})\int V_2(z-z_s)^2dz,$$

$$\delta_{0002}=\tfrac{3}{2}\int Y'dz-(\eta/\sqrt{\phi})\int V_2(z-z_s)^3dz.$$

The path method was applied by G. Wendt[6] to obtain expressions $\Delta^M x_s$ and $\Delta^M y_s$ for a single magnetic field. His equations may be obtained from Eqs. (65)–(67) by letting either $H_0$, $H_2$, and $H_4$, or $V_0$, $V_2$, and $V_4$ be identically equal to zero.

The two crossed magnetic deflection fields of actual cathode-ray tubes are ordinarily superimposed. The simultaneous deflecting action finds its expression in the existence of cross-terms with integrals of the type

$$\int X'Y'dz,\quad \int XV_2(z-z')dz,$$

$$\int YH_2(z-z_s)dz \quad \text{etc.}$$

### 3. Some Remarks Concerning Applications

The solution of the deflection problem requires a knowledge of the axial field-strength distributions $V_0(z)$ and $H_0(z)$. Experimental methods are ordinarily the only means to determine these functions. It is rarely possible to compute them from the shapes of current conductors or pole pieces. The axial field strengths $V_0$ and $H_0$ could be measured by means of a number of methods described in the literature.[7]

In order to compute the distortion and defocusing effects, two other functions are required, namely $V_2(z)$ and $H_2(z)$. From Eq. (55) it can be seen that

$$(\partial^2 H_x/\partial y^2)\big|_{\substack{x=0\\y=0}}=-2V_2(z).$$

A similar equation holds for $H_2(z)$:

$$(\partial^2 H_y/\partial x^2)\big|_{\substack{x=0\\y=0}}=2H_2(z).$$

These relations provide the means to determine $V_2$ and $H_2$ experimentally. $H_x$ is measured in the $y$, $z$ plane at $z$ for a number of values of $y$ in the neighborhood of $y=0$. It is then possible to determine for each $z$ the quantity $(\partial^2 H_x/\partial y^2)|_{x=0,\,y=0}$ and hence $V_2(z)$. A similar procedure for $H_y$ will give $H_2(z)$. The measurements are made for each field alone.

The coefficients may be determined for any screen position coordinate $z_s$. Pattern and spot distortions are computed in a manner similar to that described for electrostatic fields.

Comparing the two sets of coefficients, a remark made above can now be verified easily. The set of coefficients $\gamma_{a,b,c,d}$ goes over into the set $\delta_{b,a,d,c}$ if $H_0$ is replaced by $V_0$, $H_2$ by $V_2$ and $x_{su}$ by $y_{su}$, and $x_{su}'$ by $y_{su}'$. If the functional dependency of the field-strength functions $H_0$, $H_2$ is the same as those of $V_0$, $V_2$, corresponding coefficients will also be alike numerically and only one set of coefficients need be calculated.

# Electrostatic Dischargers for Aircraft

WAYNE C. HALL
*Naval Research Laboratory, Washington, D. C.*

(Received February 27, 1947)

Methods of discharging static electricity from aircraft in flight so as to reduce or eliminate corona interference, commonly known as precipitation static, are discussed. These methods have been classified into those employing either gaseous ions (high mobility), or those employing charge carriers such as water spray or dust particles (low mobility). Devices within the first classification offer the better solution to the problem from a practical standpoint. A discharger invented and developed at the Naval Research Laboratory is described.

## INTRODUCTION

AIRCRAFT flying in stormy weather, at a time when radio communication becomes particularly important, often encounter a severe form of radio interference sufficient to prevent all reception.[1] The interference, which is commonly known as "precipitation static," results when the aircraft, acting as an isolated conductor, becomes electrically charged. When the charging is caused by rain, snow, ice particles, or dust, that is, by friction with atmospheric particles, it is found[2] that the charging is roughly proportional to the cube of the air speed and to the frontal area of the aircraft. Consequently, this effect may be expected to increase on the larger, higher speed aircraft of the future.

The reason for precipitation static may be seen from the following. Charging the aircraft results in generating a potential gradient between the aircraft and its surroundings sufficiently large to cause a corona discharge. Corona discharges on aircraft are also caused by electrical fields existing around the aircraft in the vicinity of highly charged clouds. The corona will occur from the regions of high electrical field surrounding sharp projections on the plane, such as the antennae, antenna masts, and propellers. Ordinarily, a corona discharge creates steep wave front impulses which will induce transient voltages of considerable magnitude in the antenna circuit of a radio receiver nearby. This results because the usual corona discharge occurs in the form of bursts which are cut off in a quasi-periodic fashion by accumulations of space charge about the discharge point. In exceptional cases, as from a very sharp point, non-pulsating corona currents causing no radio interference may be observed; however, the equilibrium which permits this type of discharge is unstable and is quickly destroyed by any effect which dulls the point. If the point of discharge is on the antenna itself, a corona current as small as one microampere may cause severe static.[3]

The radio interference resulting from corona can occasionally be removed by noise suppressor circuits. Usually, however, the interference created is continuous and covers so broad a frequency range that noise reduction circuits have little or no beneficial effects. In general, precipitation static—that is, radio interference induced by corona, is best prevented by removing the cause: that is, by preventing the formation of corona on the aircraft.

An interfering corona will be prevented or reduced only if the electrical charge is drained off from the aircraft so rapidly that a potential gradient sufficient to cause corona does not develop. Evidently, any modification of the aircraft which will raise the minimum potential gradient required to cause corona will be very beneficial. It can be accomplished by the use of electrostatic shielding, or by dielectric treatment, which will tend to prevent corona discharge from the antenna or the aircraft itself. Discussion of such treatments, however, is beyond the scope of this article, which is limited to the description of discharge devices capable of removing electrostatic charge from the aircraft.

---

[1] H. M. Hucke, Proc. I.R.E. **27**, 301–316 (May, 1939); R. Gunn, W. C. Hall, and G. D. Kinzer, *ibid.* **34**, 156–161 (April, 1946).

[2] R. G. Stimmel, E. H. Rogers, F. E. Waterfall, and R. Gunn, Proc. I.R.E. **34**, 167–177 (April, 1946).

[3] G. D. Kinzer, and J. W. McGee, Proc. I.R.E. **34**, 234–240 (April, 1946).

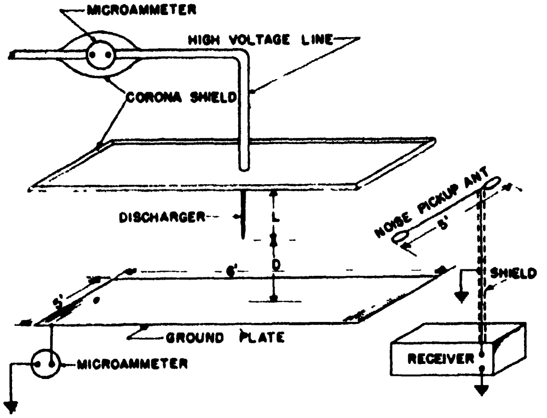


FIG. 1. Laboratory set-up to obtain discharger characteristics.

Electrostatic dischargers have the function of removing electric charge from the aircraft at a rate which substantially equals the rate of accumulation, and/or permitting the passage of electrical current to or from the aircraft when it is in the vicinity of charged clouds. Such dischargers may be classified into those using ions of high mobility (mobilities of 1.2 centimeters/second per volt/centimeter, or greater) and those using charge carriers, such as water particles or dust particles (mobilities varying from 0.002 to 0.01 centimeter/second per volt/centimeter). In the first classification are those dischargers involving the use of discharge wires or discharge points, while in the second classification are those involving the use of water spray or dust particles charged by induction or triboelectric processes.

## GASEOUS-ION DISCHARGERS

Consider, first, the static dischargers which involve the highly mobile gaseous ions and electrons. Steps involved in operating these dischargers are: (1) the production of ions, and (2) the subsequent removal of ions with one sign only from the charged aircraft by the surrounding electric field. The discharge current resulting will remove charge of the same sign as that on the aircraft, and will be limited primarily by the space charge that forms about the discharge points. An example of a discharger in this class is the so-called trailing wire discharger, which consists of a five-foot strand of small wire extending in flight from the tail of the aircraft. This discharger has had some use and does help. Another example of this class of discharger is one using needle points located on the extremities of the aircraft, or on the propellers. In these examples, the electric field causing separation of charge is fixed largely by the potential of the aircraft, a factor which may not be increased greatly without causing interfering corona on the aircraft. Hence, in an investigation of possible increased output for this class of dischargers, it is sufficient to consider only whether the discharger will benefit from more profuse generation of ions, and whether the number of dischargers to be used may be increased with beneficial effects.

For a laboratory study of the static discharger, an experimental arrangement was constructed consisting of two parallel plates, between which a d.c. potential difference of up to 60,000 volts could be established. The discharger under investigation was attached to one of these plates. Nearby, a short antenna attached to an aircraft type receiver was arranged so as to pick up the radiated interference created by the discharger; then simultaneous measurements of the potential difference between the plates, current from the discharger, and radio noise picked up by the antenna were taken. All noise measurements were made in terms of the equivalent signal required from a signal generator coupled to the

TABLE I. Characteristics of sharp point dischargers.

| Type of discharger | Distance $D$ to plate | Voltage | Discharge (micro-amperes) | Radio noise at 300 kc |
|---|---|---|---|---|
| Needle point* (sharp and clean) | 40 cm | +30,000 | 4 | Less than 3 μv |
| | | +45,000 | 11 | Less than 3 μv |
| | | +60,000 | 23 | Less than 3 μv |
| | | −30,000 | −5 | Less than 3 μv |
| | | −45,000 | −14 | Less than 3 μv |
| | | −60,000 | −26 | Less than 3 μv |
| Needle point (sharp and clean) | 2.5 cm | −45,000 | −350 | 160 μv |
| | | −57,000 | −520 | Sparkover above this |
| | | +15,000 | 30 | 16 μv |
| | | +19,000 | 80 | ~400 μv—Sparkover above this |
| Two needle points (40 cm separation parallel to plate) | 40 cm | +45,000 | 20 | Less than 3 μv |
| | | +60,000 | 38 | Less than 3 μv |
| | | −45,000 | −25 | Less than 3 μv |
| | | −60,000 | −45 | Less than 3 μv |
| Blunt point with superimposed r-f arc r-f −8.0 Mc Generator—50 watts** | 18 cm | −40,000 | −33 | Less than 3 μv |
| | | −60,000 | −68 | ~4 μv |
| | | +40,000 | 27 | <3 μv |
| | | +60,000 | 60 | ~6 μv |
| Metallicised wick*** | 18 cm | +60,000 | 68 | <3 μv |
| | | −60,000 | −74 | <3 μv |

* Using blunt points, it was observed that discharge currents were reduced and noise signals up to 400 μv were obtained.

** Estimated r-f dissipation in arc = 5 to 10 watts.

*** See Table 3. Presented here for comparison purposes.

receiver through a 100-micromicrofarad condenser to cause a receiver output, as measured by an averaging meter, equal to that from the noise signal. Measurements were taken at two frequencies, 300 kilocycles and 4000 kilocycles. Since those at 300 kilocycles were always larger, for the sake of brevity only they will be given. The laboratory set-up is shown in Fig. 1.

Using sharp needle points about 10 centimeters long, with a separation of 40 centimeters between point and plate, results were obtained which are given in Table I. When dull or damaged needles were tried, noise measurements as great as 400 microvolts were observed.[4] Next, the experiment was repeated with short lengths of wire parallel to the plate, so as to duplicate, roughly, the trailing wire discharger. The results obtained are not given, because they are comparable to those obtained with the sharp needle points, except that the radio noise was much greater. In supplementary flight experiments, it was found that if very long lengths of wire, say 400 feet in length, were trailed behind an aircraft in flight, the entire wire exhibited corona and maintained the aircraft at a low potential. Although effective in some respects, such a discharger is not practical because of the radio noise generated, the difficulty of handling such a long wire, the undesirable alteration of the radio transmission characteristics of the aircraft, and the lightning hazard involved. The short trailing wire and the blunted needle point are practical but suffer from the defect, already explained, that the corona discharge from them causes noise in nearby radio receivers. In fact, it is possible to use either type of discharger with some degree of success only when it is suitably attached to the aircraft by a series decoupling resistor of about ½ megohm or more, and when it is properly installed on the tail of the aircraft, where the interference radiated will be shielded as much as possible from the overhead antennae. This means that trailing wire dischargers or needle points (whose sharp points soon become dull or damaged) should not be installed on the wings where good inductive coupling with the antennae exists. Such dischargers, therefore, are undesirably limited in use.

TABLE II. Characteristics of dischargers with thermal ions.

| Source of thermal ions | Distance $D$ to plate | Voltage | Discharge (microamperes) | Radio noise at 300 kc** |
|---|---|---|---|---|
| Acetylene flame* (blue cone—1 cm long) | 60 cm from flame to plate | +30,000 | 6 | <3 μv |
| | | +45,000 | 16 | <3 μv |
| | | +60,000 | 28 | <3 μv |
| | | −30,000 | −12 | <3 μv |
| | | −45,000 | −26 | 3 μv |
| | | −60,000 | −44 | 4 μv |
| Gasoline blowtorch (flame length—13 cm) | 54 cm from flame to plate | +45,000 | 14 | Radio noise input less than 3 μv |
| | | +60,000 | 20 | |
| | | −45,000 | −27 | |
| | | −60,000 | −43 | |
| Manufactured gas flame from Mecca burner (flame, 15 cm long) | 45 cm | +45,000 | 19 | <3 μv |
| | | +60,000 | 35 | <3 μv |
| | | −45,000 | −28 | <3 μv |
| | | −60,000 | −53 | 4 μv |
| Signal flare (flame—10 cm long) | 60 cm | +60,000 | 10 | <3 μv |
| | | −60,000 | −20 | <3 μv |

* Flame consumes about 225 cc acetylene (N.T.P.) per minute.
** Noise measurements are uncertain and may show large values for changing electric fields.

The problem remains of devising a discharger without such limitations and of investigating whether or not profuse generation of ions will improve the output of the discharger. Taking the second part of the problem first, several possible ion sources must be reviewed. Of these, the electric arc appears to be quite promising because of the known high ionization intensity existing in the arc. Also, it is known[5] that the superposition of a high frequency field upon a d.c. field at a point can be made to prevent discharge bursts causing interference. Based on this knowledge, experiments intended to measure the effectiveness of the electric arc as a static discharger have been carried out at several laboratories, including the Naval Research Laboratory. The discharger investigated at this Laboratory consisted of a blunt tungsten point supplied with radiofrequency energy from a resonant circuit and mounted on one of the two parallel plates in the d.c. field previously described. The discharge currents for an arc of about ½ inch length, requiring some 50 watts of oscillator power at a frequency of 8.0 megacycles, are given in Table I. With the receiver tuned to 300 kilocycles, the radio interference from the arc discharge was negligible; at 5 megacycles, it was serious. Evidently, therefore, the arc frequency should be several times that of the receiver frequency

---

[4] Needles employed in these experiments were common polished steel sewing needles of various sizes, the size having negligible effect upon the results. However, even a barely visible deformation of the point was sufficient to make the discharge from the point become noisy.

[5] M. O'Day, Civil Aeronautics Adm. Tech. Dev. Report No. 27 (August, 1940).

TABLE III. Characteristics of wick* dischargers.

| Type | Distance $D$ to plate | Voltage | Discharge (microamperes) | Radio noise at 300 kc |
|---|---|---|---|---|
| Glycerol impregnated (after 25 hours flight time on aircraft) | 30 cm | +40,000 | 13 | Less than 3 $\mu$v |
| | | +60,000 | 36 | Less than 3 $\mu$v |
| | | −40,000 | −18 | Less than 3 $\mu$v |
| | | −60,000 | −48 | Less than 3 $\mu$v |
| Metallized (with silver) | 30 cm | +30,000 | 12 | Less than 3 $\mu$v |
| | | +45,000 | 25 | Less than 3 $\mu$v |
| | | +60,000 | 44 | Less than 3 $\mu$v |
| | | −30 000 | −11 | Less than 3 $\mu$v |
| | | −45,000 | −26 | Less than 3 $\mu$v |
| | | −60,000 | −50 | Less than 3 $\mu$v |
| Metallized (with silver) | 2.5 cm | +15,000 | 20 | Less than 3 $\mu$v |
| | | +30,000 | 125 | Less than 3 $\mu$v |
| | | +53,000 | 350 | Sparkover above this |
| | | −30,000 | −130 | Radio noise less than 3 $\mu$v |
| | | −45,000 | −350 | |
| | | −60,000 | −620 | |

* Each wick of cotton fibers, over-all diameter ¼ inch, length approximately 6 inches.

for noise-free operation of the discharger. Results were disappointing, in that the amount of r-f power dissipated in the arc, provided it was large enough to reduce corona static, had only a small effect on the discharger current. This was determined by varying the oscillator power over the range from 50 to 150 watts; the arc length varied from ½ to 1½ inches over this range.

Parallel experiments seeking the effect of thermal ions on discharge characteristics were conducted in the apparatus already described, using flames in place of the r-f arcs previously used. It had been observed before that a flame could be made to ignite and burn in a slipstream by using a specially designed nozzle, and by using a fuel, such as acetylene, with a high heat of combustion. In the laboratory experiments, much radio noise resulted if any metallic projections having high field intensities sufficient to cause corona were present in the flame itself; however, flames could be made to discharge with no measurable radio noise. The results obtained in the laboratory using acetylene flames, gasoline, and solid fuel are given in Table II.

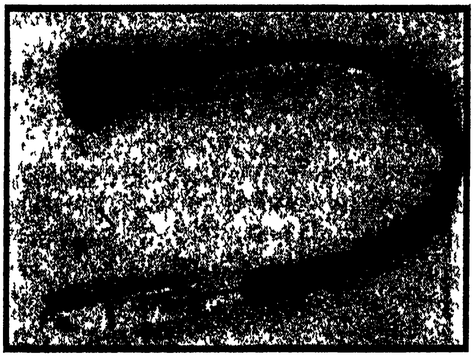


FIG. 2A. Wick type static discharger.

Next, exhaust gases from the aircraft reciprocating engines were considered for use in the discharge process; however, the low temperature, relatively speaking, of exhaust gases as they leave the exhaust pipe of this type of engine, results in ion densities too small to be of any real help. For example, calculations based on exhaust gas conductivity of eight electrostatic units (obtained from other data) indicate currents from the exhaust will be of the order of 10 to 20 microamperes per 1000 horsepower engine. This approximate result has been checked by experiment.[6] Discharge possibilities using the exhaust from jet type engines have not been investigated.

Certain observations may now be made, based on the data of Tables I and II. These data show that the point discharger, the arc discharger, and the flame discharger have roughly the same discharge current. The major part of the differences in discharge current which exist may be ascribed to differing geometries at the discharger, or to different separations between the parallel plates across which the d.c. test voltage was maintained. Neither the relatively diffuse ionized region of the flame, nor the intense ionization of the high frequency arc significantly increased the output of the discharger. The only conclusion which can be drawn is that all of these dischargers are space-charge limited and, therefore, are unavoidably limited in output.[7,8] With this

---

[6] R. G. Stimmel, E. H. Rogers, F. E. Waterfall, and R. Gunn, Proc. I.R.E. **34**, 175 (April, 1946).

[7] During discharge there is about any of the dischargers a narrow region or layer of intense ionization; the rest of the space between the plates is free from the action of the ionizing agent. Also, the discharge current to the opposite plate is only a small fraction of the maximum corresponding to the ionization. Hence, the situation is quite analogous to that discussed by J. J. Thomson, reference 8, for the case where ionization is confined to a thin layer between parallel plates. Thomson shows that in this case, the discharge current, $i$, is given by

$$i = A(V^2k/l^3)$$

where $A$ is a constant, $V$ is the potential difference between the plates, $l$ is the separation, and $k$ is the mobility of the ion carrying the charge. It is apparent that the current, $i$, is in mind, it becomes evident that improvements in the static discharger of this type can be realized only by making the discharger noise-free, cheap, easy to install, and capable of being attached at a number of points on the aircraft. The last improvement is the only one which can be expected to drain off larger and larger amounts of electric charge from the aircraft.

In the search for improvements along these lines, a "wick" discharger was developed by the author at the Naval Research Laboratory. It is essentially a small, flexible, treated cotton wick enclosed in a plastic tube, with many conducting fibers exposed at the end to serve as discharge points (see Fig. 2). The multitude of points in the wick have high resistance paths back to the aircraft, a negligible capacity, and provide a completely noise-free discharger of this type. Other fibers than cotton may be used, obviously, but none have been found more satisfactory. The cotton wick is treated to make it conducting, either by soaking the wick in an aqueous glycerol solution, chosen because it provides the desired low conductivity and has a very low rate of evaporation, or by metallizing the wicks with very thin deposits of metal. Either treatment is satisfactory, except that the glycerol treated wick is subject to a diminution in effectiveness as the glycerol evaporates or is diluted with rain water, and is more prone to become dirty in use on aircraft. The latter point is important, because wicks that are dirty may become noise. Metallized wicks using silver were found to have an indefinite life, except for mechanical wear. The discharge characteristics of some wicks made by the author are shown in Table III. These wicks generated no radio noise whatever at normal frequencies when properly made (so as to be free of metal flakes or bulky metal deposits), up to the limit of discharge which could be taken from the wicks before spark breakdown of the surrounding air occurred.

The "wick" discharger, or any discharger of this class, has limitations of the following nature. First, it must be placed on the aircraft in an exposed position where it will begin to discharge before other parts of the aircraft exhibit interfering corona. For best performance of these dischargers, therefore, they must be placed at the points of maximum electric field about the aircraft, as shown in Fig. 3. Two dischargers on each wing tip, and two to four on the empennage, represent a practical installation, though more could be used for improved performance. Second, no matter how many "wick" dischargers are used, precipitation static will occasionally be observed. This occurs when the charging current to the aircraft becomes significantly greater than the discharge current from the wicks. In such

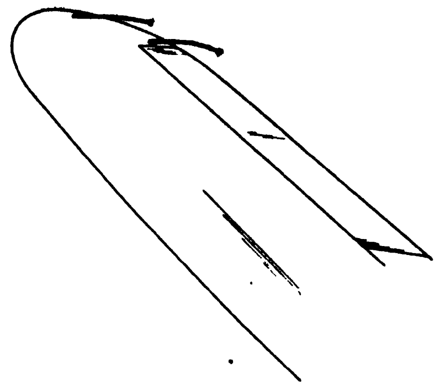


Fig. 2B. Wick type static discharger sample installation on wing.

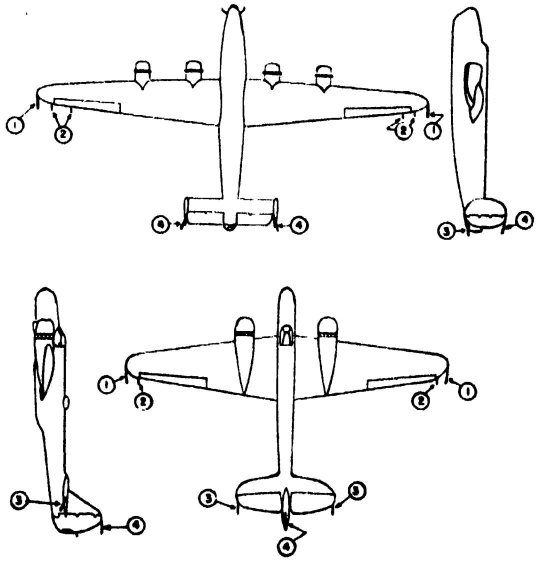


Fig. 3. Preferred localities for dischargers.

---

independent of the intensity or amount of ionization, and is thus space-charge limited.

[8] J. J. Thomson, *Conduction of Electricity through Gases* (Cambridge University Press) third edition, p. 206.

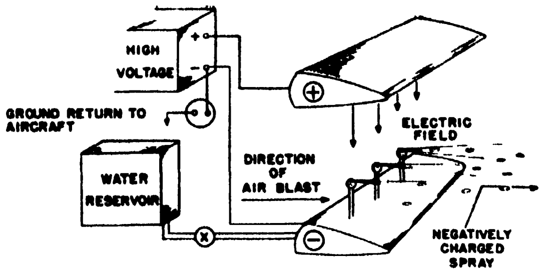


FIG. 4. Schematic of a water spray discharger.

cases, observed in severe atmospheric conditions, the aircraft will experience precipitation static, though usually for brief periods of time. For example, in the case of heavy rain, when precipitation static is heard, it will be difficult to remove. In this case, the streamers of water from the trailing edges of the airplane or other projections, such as the antenna, become, effectively, blunt points subject to corona discharge. Such streamers of water in an electric field under laboratory conditions are observed to cause intense radio static.

Another important point with regard to dischargers of this class should be made. It is known that atmospheric electrification of the aircraft may raise it to an extremely high potential with respect to its surroundings; for example, to a potential of one-half million volts. Consideration of this fact will lead to a realization that boosting the potential of a discharger relative to the aircraft by some 20,000 volts, such as may be obtained by a high voltage rectifier on the plane, will have little effect on the discharge current. Experimental results confirm this conclusion.

## HEAVY PARTICLE DISCHARGERS

The second class of electrostic dischargers involves heavy particles, such as charged water droplets, as charge carriers. Charged particles like these are produced, for example, in a liquid spray exposed to a strong electric field. Alternatively, these charge carriers are produced by a triboelectric process where certain finely divided materials, such as powdered quartz, flowing through friction tubes take on one sign of charge, leaving the other charge on the friction tube. Such particles have mobilities of the order of 0.002 centimeter/second per volt/centimeter.

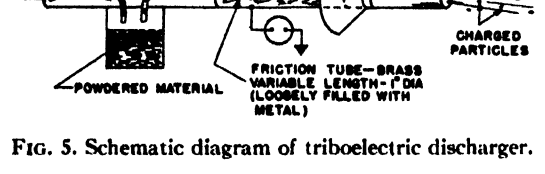


FIG. 5. Schematic diagram of triboelectric discharger.

Dischargers of this class have some advantages, in that:

(1) The external electric field about the aircraft does not necessarily affect the amount of charge induced upon the particles sprayed from the discharger; hence, the discharge currents are relatively unlimited, and the point of attachment is unimportant.

(2) The aircraft's external field does not necessarily determine the induced charge; therefore, discharge current of either sign may be obtained.

The induction type of discharger, using a water spray, is illustrated by Fig. 4. It exposes water droplets to a strong electric field as the water droplets are removed from a nozzle by a high speed blast of air. This discharger has been used very successfully to charge aircraft in flight under fair weather conditions for the purpose of simulating precipitation static conditions.[9] In flight tests using a large number of spray nozzles, discharge currents of 250 microamperes have been measured for a water discharge of approximately 75 grams/second. By stepping up the rate of flow of the water through the spray nozzle, this discharger could be made to discharge quite an appreciable current. If the auxiliary electric field that is established at the spray nozzle to charge the droplets is maintained slightly below a value sufficient to cause corona, there will be little radio noise generated. There are engineering drawbacks to such a discharger, however; for example, the discharger requires an auxiliary high voltage of some 20,000 volts, a bulky and heavy tank of non-freezing liquid such as alcohol, and it increases the aerodynamic losses of the aircraft.

The triboelectric discharger, on the other hand, is illustrated in one form that it takes in Fig. 5. Here the friction between the tube and particles moving through affects the amount of charge carried away by the particle. Since the friction

---

[9] R. C. Waddel, R. C. Drutowski, and W. N. Blatt, Proc. I.R.E. 34, 161–166 (April, 1946).

is partially determined by the velocity of the air blast through the tube, the data to be reported on the triboelectric discharger are for one air velocity only, namely, 250 miles per hour at the inlet tube. Table IV has been prepared to show the performance characteristics of this discharger under the given conditions. As high as 84 microcoulombs of charge per gram of material was obtained when using finely divided quartz (Santocel). Making the reasonable assumption that the average radius of the particles of the diatomaceous earth used (largely silica) was $10^{-4}$ centimeter, the theoretical limit of charge which could be carried by the powdered material per cubic centimeter of solid material (before pulverization) becomes $3E/4\pi R$, where $E$ is the electric field at breakdown of the surrounding air. This corresponds roughly to a figure of 80 microcoulombs per cubic centimeter, or 30 microcoulombs per gram.

With this type of discharge, the substance of the powdered material usually determines whether the sign of charge will be positive or negative. Many materials, when blown through the friction tube, came out with a negative sign. This would be desirable, because in nearly all cases of charging by atmospheric particles, the airplane became negatively charged and would require a discharger of negative electricity. Noise measurements on the triboelectric discharger gave erratic results; at times a high level of noise intensity was observed. These and other drawbacks, of an engineering and practical nature, are believed to be sufficiently important to prevent any commercial use of this discharger.

TABLE IV. Characteristics of triboelectric dischargers using powdered materials.

| Material | Discharge current (micro-amperes) | Discharge of material (grams/second) | Remarks | Efficiency micro-coulombs per gram of material |
|---|---|---|---|---|
| Colloidal carbon* | +7# | 1.4 | Not dried | 5 |
| Infusorial earth* | −14# | 1.4 | Not dried | 10 |
| Diatomaceous earth* | −18 | 1.7 | Not dried | 10.6 |
| Diatomaceous earth* | −13 | 1.1 | Dried 2½ hrs. at 200°C | 12 |
| Diatomaceous earth** | −18 | 1.4 | Dried 2½ hrs. at 200°C | 13 |
| Santocel (powdered quartz from Monsanto Chemical Co.)** | −50 | 1.0 | Not dried | 50 |
| Santocel (powdered quartz from Monsanto Chemical Co.)*** | −33 | 0.4 | Not dried | 84 |
| Diatomaceous earth*** | −12 | 0.36 | Dried 2½ hrs. at 200°C | 33 |
| Santocel**** | −9.5 | 0.13 | Not dried | 73 |
| Santocel**** | −25 | 0.30 | Dried 1½ hrs. at 180°C | 83 |
| Titanium dioxide**** | −5.5 | 0.60 | Not dried | 9 |
| Graphite**** | 0 | 1.0 | Not dried | 0 |
| Wheat flour**** | +2.5 | 0.4 | Not dried | 7 |

# Denotes sign of charge on powder.

* Brass friction tube encloses ten copper tubes, ½ cm diam., 30 cm long.

** Brass friction tube, 75 cm long, loosely filled with steel wool.

*** Brass friction tube, 245 cm long, loosely filled with copper turnings.

**** Brass friction tube, 40 cm long, loosely filled with copper screen, mesh = 16/inch.

## SUMMARY

To summarize, it appears that static dischargers are a practical necessity on aircraft using present-day methods of radio communication. Wicks or trailing wires, being simple and having relatively good performance, are to be preferred over other types. The wick discharger, if properly made, is completely noise-free and is applicable to the airplane at a number of points; hence, it appears to be the best available discharger. No discharger, alone, however, offers a complete solution to the problem.

# Slow Transverse Magnetic Waves in Cylindrical Guides

G. G. Bruck and E. R. Wicher
*Specialties, Inc., Syosset, Long Island, New York*

(Received March 24, 1947)

The fundamental physical phenomenon upon which linear electron accelerators and traveling beam tubes depend is the fact that the phase velocity of guided *TM* waves can be reduced to a fraction of the velocity of light. There are various practical ways of achieving these reduced velocities, such as diaphragms or spiral grooves in the guide wall. An open helix has been used by Pierce.

An accurate knowledge of the field pattern would greatly facilitate an understanding of these devices. It is of particular importance to have exact information about the field in the region in which the charge actually travels, namely, in a region located at least a considerable fraction of a wave-length from the guide walls. Because of the complex shapes assumed by actual guide walls calculations of the field are inevitably somewhat inexact and complicated.

The authors propose that, for the purposes of theory, the well-established proposition concerning the equivalence of true and simulated dielectrics in producing a slow field be used. We propose to think of slowing down the phase velocity by lining the guide walls with a natural dielectric. The essential properties of the slow field, particularly in the important region referred to above, will not be affected by this method of obtaining it while, on the other hand, the field calculation becomes quite easy and exact.

Although this procedure is intended in the main as a device to facilitate the theoretical calculation of the field, the authors believe that such a step—replacement of periodic metallic structures by dielectrics—might prove useful in some actual applications.

---

The solutions of Maxwell's equations appropriate to the description of a $TM_{01}$ mode propagated axially in a cylindrical tube, may be written

$$
\begin{aligned}
&E_z = AJ_0(\kappa\rho)+BN_0(\kappa\rho)\\
&E_r = (2\pi i/\kappa s)[AJ_1(\kappa\rho)+BN_1(\kappa\rho)]\\
&E_\varphi=0,\ H_z=0,\ H_r=0\\
&H_\varphi=(2\pi ic\epsilon/\kappa)[AJ_1(\kappa\rho)+BN_1(\kappa\rho)]
\end{aligned}
\tag{1}
$$

where $\rho=r/\lambda$, $s=v/c$, $\lambda$ = free space wave-length, $c$ = velocity of light, $v$ = phase velocity actually present in tube, $(z, r, \varphi)$ are standard cylindrical coordinates, $\kappa=2\pi c(\mu\epsilon-1/v^2)^{\frac{1}{2}}$, $\epsilon$ = inductive capacity of medium, $\mu$ = permeability, $(E_z, E_r, E_\varphi)$, are components of electric field strength, $(H_z, H_r, H_\varphi)$, are components of magnetic field strength, $A$, $B$ are arbitrary constants to be adjusted to fit assigned boundary conditions.

Georgi rationalized units are used. It is understood that the factor $\exp(i\omega t-\gamma z)$ has been omitted from the right members of all of (1). Here $\omega$ is the angular frequency, equal to $2\pi c/\lambda$, and $\gamma$ is the propagation constant. We neglect conductivity so that $\gamma=i\omega/v=2\pi i/s\lambda$.

Table I.

| $\rho_2$ | $\rho_2-\rho_1$ | $\epsilon_2/\epsilon_1$ |
|---|---|---|
| 0.9564 | 0.00064 | 154000.0 |
| 0.05530 | 0.00530 | 2499.0 |
| 0.05943 | 0.00943 | 774.4 |
| 0.06164 | 0.01164 | 524.3 |
| 0.07187 | 0.02187 | 174.4 |
| 0.08958 | 0.03958 | 74.44 |
| 0.10000 | 0.05000 | 57.25 |
| 0.1500 | 0.1000 | 34.27 |
| 0.4897 | 0.4397 | 25.63 |

We apply the solutions (1) to the case of a pair of infinitely long coaxial cylinders with radii $r_1=\rho_1\lambda$ and $r_2=\rho_2\lambda$, $\rho_1<\rho_2$. The inner cylinder is evacuated and the region $\rho_1<\rho<\rho_2$ is filled with a dielectric of inductive capacity $\epsilon_2$. The outer cylinder is bounded by a perfect conductor. In the following we call the region $\rho<\rho_1$, region I, and the region $\rho_1<\rho<\rho_2$, region II. Parameters referring explicitly to these regions are distinguished by the use of subscripts 1 and 2.

In medium *I*, $B_1=0$ because the *N* functions are infinite at $\rho=0$. Hence,

$$
\begin{aligned}
&E_{z1}=A_1J_0(\kappa_1\rho),\quad E_{r1}=(2\pi i/\kappa_1 s)A_1J_1(\kappa_1\rho),\\
&E_{\varphi1}=0,\quad H_{z1}=0,\quad H_{r1}=0,\\
&H_{\varphi1}=(2\pi ic\epsilon_1/\kappa_1)A_1J_1(\kappa\rho),
\end{aligned}
\tag{2}
$$

$$\kappa_1=2\pi(1-1/s^2)^{\frac{1}{2}}. \tag{3}$$

We now examine three cases.

Case *I*. $\kappa_1$ is real. Then by (3), $s>1$ and $v>c$. The $z$ component of the electric field is described by the function $J_0(\kappa_1\rho)$ which is unity at $\rho=0$, and which tapers off to a zero at $\kappa_1\rho=2.4048$. In general, the fact that $J_0(\kappa_1\rho)$ is a decreasing

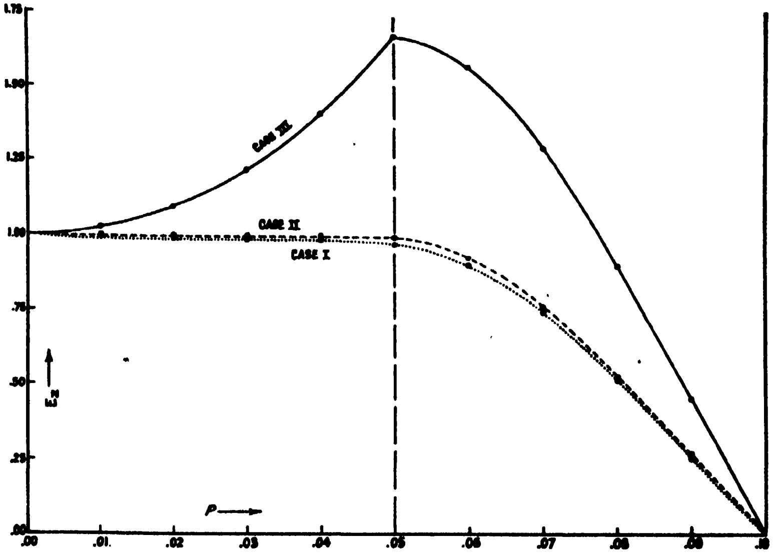


FIG. 1. Distribution of the longitudinal component of electric field strength through a cross section of the tube.

function for real $\kappa_1$ makes it usable in any case where $v>c$.

Case *II*. $\kappa_1$ is zero. Then by (3), $s=1$ and $v=c$. This gives $E_{z1}=A_1$, a constant. If region *I* were bounded by a conductor, this result would be trivial since all field components would have to vanish. Here, the result will have application to cases of electron accelerators, or traveling wave tubes in which the electrons have an energy of more than $10^6$ electron volts.

Case *III*. $\kappa_1$ is imaginary. In this case $s<1$ and $v>c$. This gives an $E_{z1}$ distribution described by $J_0(i|\kappa_1|\rho)=I_0(|\kappa_1|\rho)$, an increasing function of $\rho$. This type of distribution is of particular interest because it permits slowing down the phase velocity of the field to any desired value.

In region *II* the fields will be given by

$$\begin{aligned}
&E_{z2}=A_2J_0(\kappa_2\rho)+B_2N_0(\kappa_2\rho)\\
&E_{r2}=(2\pi i/\kappa_2 s)[A_2J_1(\kappa_2\rho)+B_2N_1(\kappa_2\rho)] \qquad (4)\\
&E_{\varphi 2}=0,\ H_{z2}=0,\ H_{r2}=0
\end{aligned}$$

$$\left.\begin{aligned}
&H_{\varphi 2}=(2\pi ic\epsilon_2/\kappa_2)[A_2J_1(\kappa_2\rho)+B_2N_1(\kappa_2\rho)]\\
&\kappa_2=2\pi[(\epsilon_2\mu_2/\epsilon_1\mu_1)-(1/s^2)]^{\frac{1}{2}}.
\end{aligned}\right\} \quad (5)$$

The boundary conditions may be written

$$\begin{aligned}
&E_{z2}(\kappa_2\rho_2)=0,\\
&E_{z1}(\kappa_1\rho_1)=E_{z2}(\kappa_2\rho_1), \qquad (6)\\
&\epsilon_1E_{r1}(\kappa_1\rho_1)=\epsilon_2E_{r2}(\kappa_2\rho_1).
\end{aligned}$$

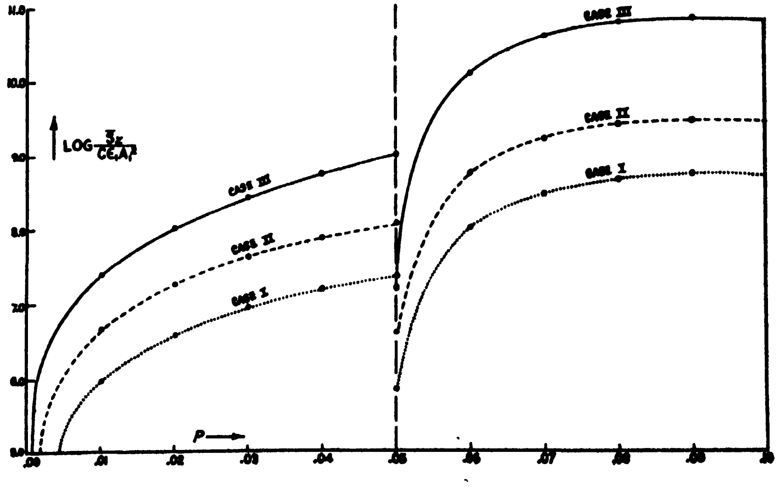


FIG. 2. Energy distribution. The logarithm of the real part of the complex Poynting vector plotted against the radial coordinate of the tube. The logarithm has been used to obtain a convenient scale.

From (2), (4), and (6) we get

$$A_2/A_1=(\pi\kappa_2\rho_1/2)\big[(\epsilon_1/\epsilon_2)(\kappa_2/\kappa_1)N_0(\kappa_2\rho_1)J_1(\kappa_1\rho_1)-N_1(\kappa_2\rho_1)J_0(\kappa_1\rho_1)\big],$$
$$B_2/A_1=(\pi\kappa_2\rho_1/2)\big[J_0(\kappa_1\rho_1)J_1(\kappa_2\rho_1)-(\epsilon_1/\epsilon_2)(\kappa_2/\kappa_1)J_0(\kappa_2\rho_1)J_1(\kappa_1\rho_1)\big]. \tag{7}$$

In these formulas $A_1$ is simply the amplitude of the $z$ component of the electric field along the axis. From (6) and (7) we obtain

$$\frac{\epsilon_1}{\epsilon_2}\frac{\kappa_2}{\kappa_1}\frac{J_1(\kappa_1\rho_1)}{J_0(\kappa_1\rho_1)}=\frac{N_0(\kappa_2\rho_2)J_1(\kappa_2\rho_1)-J_0(\kappa_2\rho_2)N_1(\kappa_2\rho_1)}{N_0(\kappa_2\rho_2)J_0(\kappa_2\rho_1)-J_0(\kappa_2\rho_2)N_0(\kappa_2\rho_1)}. \tag{8}$$

This formula, taken with the definitions of $\kappa_1$ and $\kappa_2$, is a relation between field velocity, tube radii, and the dielectric constant of region *II*. In order to have a definite Case *III* to discuss, we chose $s=v/c=0.2$ and let $\rho_1=0.05$. These choices are arbitrary, and are simply meant to be illustrative. Having made this selection, (8) can be solved by a combination of patience with a variety of numerical procedures. The result is shown in Table I.

One particularly useful device in obtaining these points is to select the product $\kappa_2\rho_2$ to be a zero of either $J_0(\kappa_2\rho_2)$ or $N_0(\kappa_2\rho_2)$. For general points, (8) may be transformed to read

$$\frac{\epsilon_1}{\epsilon_2}\frac{\kappa_2}{\kappa_1}\frac{J_1(\kappa_1\rho_1)}{J_0(\kappa_1\rho_1)}=\frac{|H_1^{(1)}(\kappa_2\rho_1)|\sin(\psi-\theta_1)}{|H_0^{(1)}(\kappa_2\rho_1)|\sin(\psi-\theta_0)}, \tag{9}$$

where $\psi=\arg H_0^{(1)}(\kappa_2\rho_2)$, $\theta=\arg H_0^{(1)}(\kappa_2\rho_1)$, $\theta_1=\arg H_1^{(1)}(\kappa_2\rho_1)$.

The tables in Watson† are simply applicable to this form. We did the calculations from (9) to four significant figures and checked the results by substitution into (8). It will be noticed that for a given $\epsilon_2/\epsilon_1$, a denumerable set of numbers $\rho_2$ exists; we tabulate only the lowest $\rho_2$.

To compare field strength distributions, energy distributions, and field patterns in the three cases, we selected $v=0.2c$ (Case *III*), $v=c$ (Case *II*), and $v=5c$ (Case *I*), and fixed $\rho_1=0.05$ and

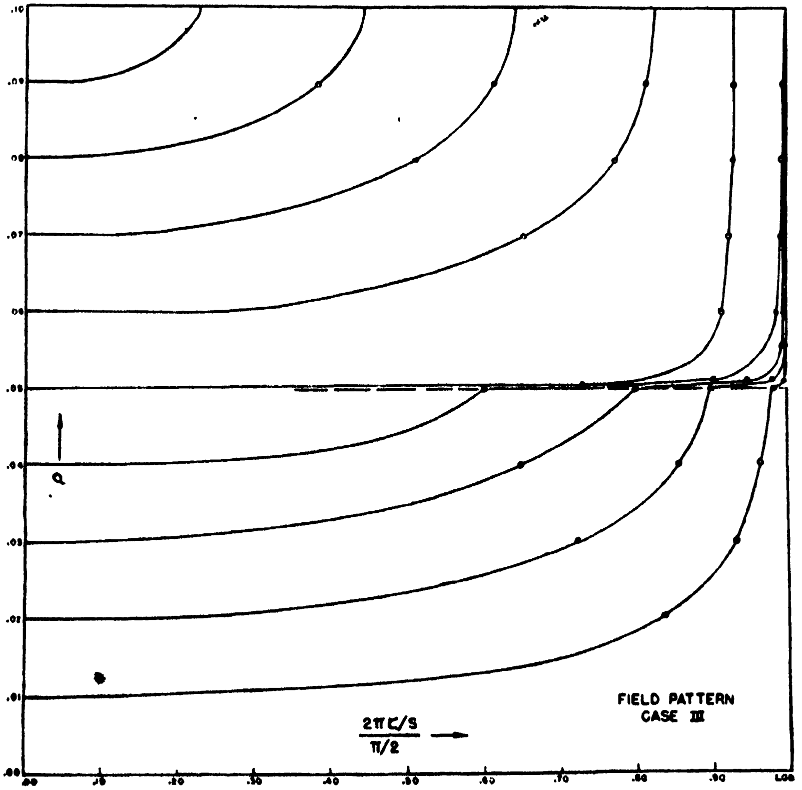


FIG. 3. Longitudinal field pattern, showing lines connecting points of equal electric field strength along a longitudinal cross section of the tube. The slope of the lines gives the direction of the field. $\zeta=z/\lambda$, and $s=v/c=2/10$.

---

† G. N. Watson, *Theory of Bessel Functions* (The Macmillan Company, New York, 1945), Chapter XX.

$\rho=0.10$ in all cases. This gives $\epsilon_2/\epsilon_1=57.25$ (Case *III*), $\epsilon_2/\epsilon_1=32.80$ (Case *II*), $\epsilon_2/\epsilon_1=31.82$ (Case *I*).

The results of this comparison for the distribution of the $z$ component of the electric field are seen in Fig. 1.

The real part of the complex Poynting vector is given by

$$\bar{S}_{z1}/c\epsilon_1 A_1^2=(2\pi^2/\kappa_1^2 s)[J_1(\kappa_1\rho)]^2,$$

$$\bar{S}_{z2}/c\epsilon_1 A_1^2=\frac{2\pi^2}{\kappa_2^2 s}\frac{\epsilon_2}{\epsilon_1}\left[\frac{A_2}{A_1}J_1(\kappa_2\rho)+\frac{B_2}{A_1}N_1(\kappa_2\rho)\right]^2. \tag{10}$$

Figure 2 shows the logarithm of this quantity as a function of $\rho$ for the three cases. It is obvious that practically all the energy is carried in region *II* in each case.

The field pattern is given by integrating the relation

$$dz/dr=ReE_z/ReE_r. \tag{11}$$

Letting

$$\zeta=z/\lambda, \tag{12}$$

(11) yields

$$\begin{aligned}\kappa_1\rho J_1(\kappa_1\rho)&=C_1\sec(2\pi\zeta/s)\\ \kappa_2\rho[A_2J_1(\kappa_2\rho)+B_2N_1(\kappa_2\rho)]&=C_2\sec(2\pi\zeta/s),\end{aligned} \tag{13}$$

where $C_1$ and $C_2$ are integration constants. The obvious requirement on these constants is

$$\frac{C_1}{C_2}=\frac{\kappa_1\rho_1J_1(\kappa_1\rho_1)}{\kappa_2\rho_1[A_2J_1(\kappa_2\rho_1)+B_2N_1(\kappa_2\rho_1)]}. \tag{14}$$

In Case *II* the first of Eqs. (13) may be replaced by

$$\rho^2=C_1\sec 2\pi\zeta, \tag{15}$$

so that (14) becomes, for this case

$$\frac{C_1}{C_2}=\frac{\rho_1^2}{\kappa_2\rho_1[A_2J_1(\kappa_2\rho_1)+B_2N_1(\kappa_2\rho_1)]}. \tag{16}$$

It develops that if the pattern be drawn by plotting $\rho$ against $2\pi\zeta/s$, there is no appreciable difference between the three cases. This means that the effect of decreasing the field velocity is simply to compress the pattern in the direction of propagation. Figure 3 shows the pattern for Case *III*.

---

# Incandescent Lamps as Electrical Load for Testing

S. Keilien
*Pass and Seymour, Inc., Syracuse, New York*

(Received March 25, 1947)

Because of the positive coefficient of resistivity of tungsten, incandescent lamps permit a current considerably greater than the normal lamp current to flow for a short initial period. When lamps are used as load for testing electrical apparatus, the factors affecting this inrush must be considered. Size of power source, its distance from the load (lamps), and frequency of switching the lamps (employing repetitive operations for endurance testing) are evaluated.

From the given curve the necessary cooling time for the desired current inrush may be determined, and thus the frequency with which the lamps could be switched "on" and "off" may be calculated. Practical means for obtaining the required cooling time are described.

ONE of an electric control device's usual function is switching. In this operation the electrical load is either connected to or disconnected from the source of power supply and, to be acceptable, the switching element should be capable of performing with least deterioration.

There are 3 basic types of loads, and their characteristics affecting the performance of switching devices may be summed up as follows:

Non-inductive-resistance load-current rise to reach the Ohm's-law value and its fall to zero (when disconnected) is instantaneous; the values of steady-state conditions are not exceeded.

Inductive load-current rise is gradual, depending on the rate of magnetic flux build-up; on disconnection of this load the stored electromagnetic energy is dissipated quickly, causing momentarily a high voltage, frequently far above that of the supply.

Tungsten-filament lamp load-current rise is very rapid and the maximum exceeds that of the steady-state value; current decay to normal value, following its reaching the peak, is relatively slow; upon the disconnection the fall of current to zero is instantaneous.

TABLE I.* Ratio *R* for various size lamps.

| Size of lamp watts | Ratio *R* |
|---|---|
| 100 | 9.1 |
| 500 | 8.7 |
| 1000 | 6.8 |

* R. W. Gillette, "Control of Multiple Street Lighting," Elec. World **94**, 419 (1929).

From the above grouping it will be seen that the inductive and the tungsten-filament lamp loads call for special precautions to be taken: the former upon interruption of the circuit (because of presence of high voltage) and the latter at completion of the circuit (because of current inrush).

In practice, the usual non-inductive-resistance loads are heating appliances of various kinds; the inductive load is typified by magnets and motors; the lamp-load characteristics are present in tungsten-filament lamps and condensers. For determination of suitability of a switching device, it is usually tested under various load conditions, and it is important to give proper consideration to these loads if the results of tests are to be used as criteria.

Tungsten filament lamps are employed extensively for testing the devices which are designed for such loads, or which are intended for loads with initial inrush of current. It is important, therefore, to have a clear conception of lamp characteristics and of limitations to assure their proper application as loads for testing.

TABLE II. Ratio *R* for various size lamp loads and their distance from busbars.

| Load connecting circuit | | Load (watts) and number and size of lamps (w) | | | |
|---|---|---|---|---|---|
| Length (ft.) and wire size (B and S gauge) | Total wire res. ohms | 2400 w<br>4 500 w<br>2-200 w | 1200 w<br>2 500 w<br>1-200 w | 520 w<br>8 60 w<br>1 40 w | 500 w<br>1 500 w |
| + 163′—No. 10<br>16′—No. 14 | .203 | — | – | 9.4 | – |
| + 148′—No. 10<br>16′—No. 14 | .188 | — | — | 9.6 | – |
| 163′—No. 10 | .163 | 7.7 | 9.1 | — | 11.0 |
| 148′—No. 10 | .148 | 8.1 | 9.3 | — | 11.2 |
| 10′—No. 10<br>+ 10′—No. 11<br>+ 16′—No. 14 | .063 | - | — | 10.9 | — |
| + 10′—No. 10<br>10′—No. 11 | .023 | 8.9 | 10.3 | - | 11.4 |

Notes: 1. All lamps—115 volts.
2. Power derived from specially designed 125 volts d.c. generator.
3. Values of "length" in first column—length of conductors between load and busbars.
4. Plus sign (+) in first column indicates series connections.

Because of the positive temperature coefficient of resistivity of the metal, the tungsten-filament lamps possess low resistance when cold and higher resistance when heated. It must be noted that carbon-filament lamps have the reverse characteristics because of the negative coefficient of resistivity of carbon. For regular 115- or 120-volt tungsten-filament lamps, sizes 40 to 1500 watts, the operating temperature of the filament varies between 4400°F and 4900°F, and the resistance of a hot lamp is 15 to 17 times that of a cold lamp. Consequently, if a cold lamp is connected to a source of supply, a current considerably greater than the normal lamp current will flow for a short time. Because of heating by the passing current, the temperature of the filament will rise and so will its resistance, thus causing a corresponding decrease in current. It would appear that the ratio of inrush current to the normal lamp current would be also 15 or 17 to 1. However, these values are not attained usually, for the inrush current also depends on such factors as the capacity of the current source, impedance of the circuit, and voltage stability, all of these tending to reduce the magnitude of the peak.

It is difficult to determine quantitatively the effect of some of the above factors, but definite conclusions may be drawn on the basis of experimental data.

The source of power should be of ample capacity to furnish the necessary inrush current. It is apparent that the source should not be the limiting factor, since the characteristics of the filament determine these inrush values. The ratio *R* of the peak current to nominal current of the lamp may be used as the criterion in evaluating the suitability of the generator.

A small generator having a relatively slow current response and a small amount of available power will give a smaller value of *R*, for the same lamp load, than would a larger machine. A 5-kw, 125 volt, direct-current generator controlling tungsten-filament lamps gave the current ratios *R* shown in Table I. A generator used for routine and experimental testing of switches produced considerably greater values of $R$, as shown in Table II. It will be observed that such generator permits a current inrush to produce a value of $R$ as high as 10.3, for 1200-watt load (compare a corresponding value of 6.8 for 1000-watt load with a smaller generator).

A generator for testing switching devices with lamps as load must be properly designed, for its rated capacity is not the only consideration. Factors like armature reaction, commutation, and voltage stability for a specific load will affect the values of $R$. The Underwriters' Laboratories, Inc., standard for switches specifies a generator capable of producing a current inrush of eight times the normal value of current, using a 20-ampere (normal) lamp load. Values of $R$ considerably in excess of this can be obtained with specially designed generators.[1]

Size of conductors and distance between the lamp and the generator affect the resistance of the circuit. Since the current inrush is of transient nature, the impedance of this circuit must also be considered. The last four columns of Table II give values of the ratio $R$ for various conductor sizes and their lengths. From this table it may also be observed that, for the same wattage, the larger the number of lamps in the circuit, the lower is the value of $R$ (compare last two columns). This may be explained as follows: The rate of heating of a small lamp is considerably higher than that of a large lamp.[2] Because of this, with smaller lamps there will be considerably greater overlapping of the initial period when the current is limited by the circuit inductance and of the period when resistance of filament has risen sufficiently to reduce the current. Therefore, with same current, inductance and resistance of the circuit, smaller lamps would produce a lower value of $R$ than would one large lamp of the same wattage. The $R$ in the case of smaller lamps is further diminished by their lower final filament temperature, as compared with that of the larger lamp.

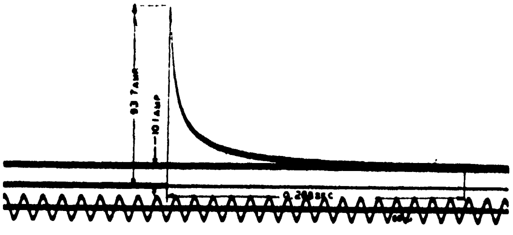


FIG. 1. Oscillogram showing current rise and decay in 1200-watt lamp circuit (2–500-w and 1–200-w lamps) upon its connection to 125-volt d.c. generator.

The use of tungsten-filament lamps for repetitive operations (connecting to and disconnecting from power source) should be conducted with discretion because of the time required to bring the filament to final temperature and the time needed for cooling it to the ambient temperature. The filament-heating time is relatively short. From oscillogram, Fig. 1, it will be seen that normal current was reached within 0.268 second. For lamps of lower rating this time is correspondingly shorter. Since the current decay in the lamp after reaching inrush peak is exponential, it is difficult to determine with a great degree of precision the exact point where the current curve reaches its normal value. However, for practical purposes, the oscillographic determination is sufficiently accurate.

The rate of cooling of the lamp is important in the calculation of the frequency with which the device under test may be operated. The value of $R$ (ratio of inrush peak current to normal lamp current) may be used as basis for the determination. Should a lamp be reconnected to the power source while the filament is still hot, the ratio $R$ would not be as great as that for a completely cooled lamp. This is apparent,

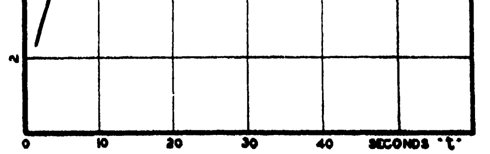


FIG. 2. Relation between cooling time $t$ of tungsten-filament lamps and ratio $R$ (ratio of initial peak current to normal current) for 1200-watt lamp load at 125-v d.c.

---

[1] S. Keilien, "Improving products through research and test for quality," Product Engineering 10, 24 (1939).

[2] W. E. Forsythe, M. A. Easley, and D. D. Hinman, "Time constants of incandescent lamps," J. App. Phys. 9, 209 (1938).

for under such a condition the differential in temperature would be lower. Results of experiments conducted with lamps which were cooled for varying intervals are represented by a curve, Fig. 2. In order to obtain stable data the readings of current inrush (from oscillograms) were obtained after constant ambient temperature had been established.

The curve is particularly valuable when rhythmic operation of switching *on* and *off* of lamps is to be employed. In testing current-interrupting devices for endurance, it is important to have a specific current inrush. From Fig. 2, it is possible to determine the cooling time (duration of the *off* period) for the desired current inrush, and then the frequency with which the testing is to be conducted may be readily calculated.

If, for endurance testing, only one lamp (or one bank of lamps) were used, it would be necessary to allow considerable time for its cooling in order to obtain the desired current peak on the subsequent *on* operation. Such procedure would be quite lengthy and would impede testing, especially if a large number of specimens is to be tested to keep up with production schedules.

To facilitate testing, a number of lamps (or banks of lamps) may be employed, so arranged that while one is being used (*on*) the others are cooling (*off*). This method makes possible any desired rate of specimen operation, the limitations being only those of the specimen itself, such as heat dissipation of its current carrying parts, velocity of its moving system, etc., and the time required to bring the filament to maximum temperature.

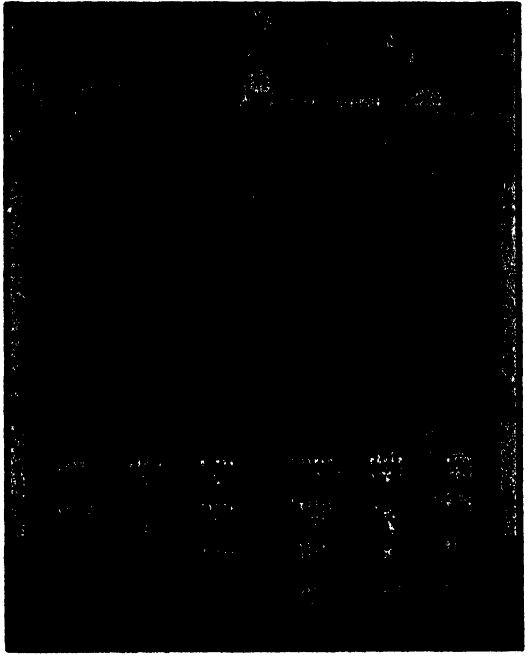


FIG. 3. View of switchboard with equipment for automatic selection of lamp circuits.

For repetitive operation of switches at a uniform rate, the number of lamps required to produce a desired cooling time may be calculated from Eq. (1),

$$N=\frac{t+a}{b+a} \tag{1}$$

where:

$N$—number of lamps (or banks of lamps),

$t$—lamp-cooling time to produce the desired value of $R$ in seconds (obtainable from curve Fig. 2, for specific load),

$a$—time, per cycle of operation, during which the switch is *on*, in seconds,

$b$—time, per cycle of operation, during which the switch is *off*, in seconds.

If the *on* and the *off* periods of switch operation are equal, then Eq. (1) becomes

$$N=\frac{1}{60}(t\times f+30) \tag{2}$$

where $f$ = rate of switch operation, in cycles per minute.

In testing switches, $a$ is usually greater than the time required to heat the lamp and since the heating time is quite negligible in comparison with that needed for cooling the lamp, it was not considered in the formulae.

A very satisfactory scheme for the simultaneous testing of a number of switches is to connect the cold lamp in the circuit containing the specimen under test while this specimen is *off*, and to disconnect the lamp while the specimen is in the succeeding *off* position. Thus, the lamp-selecting device does not ***make*** nor ***break*** the circuit while it is alive; it is done by the device undergoing the test. The actual equipment for accomplishing such selection is shown in Fig. 3, which is a part of an installation for simultaneous testing of 10 switches, each with a 1200-watt lamp load, or a smaller number of switches with a proportionately greater load.

The cool lamps are connected in the test circuits by means of a magnetically-operated, step-by-step selector, which is driven in synchronism with the switches under test. The selector actually controls the coils of heavy-duty contactors, the contacts of which introduce into or remove from circuits the lamps for loading the switches. Other means of selection of cool lamps, such as continuously or intermittently-operated commutators, or cam-operated contacts of either spring or mercury type may be employed.

In some cases it is desired to obtain the inrush current with lesser equipment, or at quite rapid rate. Then, it is possible to connect resistance loads in such a way that the low resistance (allowing high current inrush) is removed from the circuit a short time after the switch under test has established the contact, leaving a higher resistance in the circuit (corresponding to normal current of lamp), thus permitting only normal current interruption by the switch. This arrangement does not duplicate actual lamp operation, because if current equivalent to inrush value is *on* for a considerable time the contacts are subjected to excessive heating. Furthermore, in the event of contact vibrations the contact erosion and welding are aggravated. Consequently, such testing may become unduly severe. In this scheme, shown in Fig. 4, the normal current is determined by resistance $R_a$, while the inrush current is limited by resistance $R_b$. After a predetermined interval following the closure of switch $S$, the timing contacts $T$ open, thus permitting the switch $S$ to break only the normal current, as determined by $R_a$. The current rises and decays abruptly since non-inductive resistances are used. Contacts $T$ are synchronized with $S$ so that they *close* and *open* before $S$ is subjected to respective operations.

Another method is based on the characteristic curve of charging a condenser. A condenser of suitable capacity is connected in the test circuit before the switch is *closed;* this produces the necessary current inrush and decay. The condenser is then disconnected by a relay after a desired lapse of time, restoring the circuit containing a resistance to a current equal to that of normal value of the lamp. Essentially this arrangement is a modification of that shown in Fig. 4, the difference being in the introduction of capacitance $C$ (not shown in diagram) in series with resistance $R_b$. As in the previous scheme, the resistance $R_a$ determines the final value of current flowing through switch (normal lamp current), resistance $R_b$ determines the peak current, while condenser $C$ defines the rate of current decay during the transient period upon closure of switch $S$. Contacts $T$ disconnect the condenser $C$ (and resistance $R_b$) when the current through this path has decayed to a desired magnitude (near zero). This is done before $S$ is opened. The operation of the timing contacts $T$ may be accomplished either electrically or mechanically. Provision must be made for discharging the condenser $C$ prior to its reconnecting into the circuit.

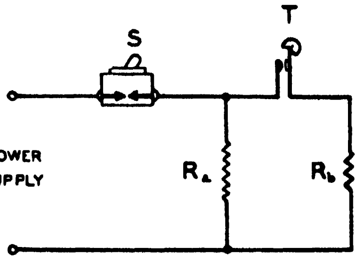


FIG. 4. Schematic wiring diagram of connections for producing initial current inrush.

Because of presence of some inductance in the circuit, it is difficult to predict by calculation the initial peak current that would result. It is, therefore, necessary to adjust the parameter of the circuit until the desired values are obtained. These constants will differ from the calculated values.

While it is possible to duplicate relatively closely a current rise and decay of a tungsten-filament lamp, such synthetic-load testing equipment has been found less desirable from the standpoint of stability, reliability, and maintenance for large scale testing.

The tungsten-filament lamps should be employed as electrical load with due consideration given to their characteristics. Particular attention should be given to the frequency of operation of specimens under test, in order to obtain the desired current inrush during the *on* period. The source of power supply and the impedance of the circuit must also be evaluated if it is planned to duplicate the conditions to which the switching device would be subjected under actual field operations.

# The Capacity per Unit Length and Characteristic Impedance of Coaxial Cables with One Slightly Non-Circular Conductor

PHILIP PARZEN*
*New York, New York*

(Received April 9, 1947)

It is possible to calculate approximately the electrostatic field in a coaxial line with one slightly non-circular conductor as a perturbation of that existing in the usual coaxial line. General expressions are then derived for the capacity per unit length and the characteristic impedance of such lines. In particular, these results are applied to the coaxial line with outer square conductor.

## INTRODUCTION

THE calculation of the capacity per unit length, $C$, of a coaxial line depends upon the derivation of a potential distribution function, $V$, which satisfies $\nabla^2 V=0$, with the boundary conditions: $V=0$ on inner conductor, $V=V_1$ =constant on outer conductor. Thus,

$$C=-K\int \frac{\partial V}{\partial n} ds/V_1 \quad \text{(m.k.s. units)} \qquad (1)$$

where the numerator is the total charge per unit length which is given by the line integral of the normal derivative of the potential along either conductor, and $K$ is the dielectric constant of the medium.

The characteristic impedance, $Z_0$, of a resistanceless line is:

$$Z_0=1/Cv,\ v=\text{velocity of light in medium.} \qquad (2)$$

This may be seen physically as follows: an outgoing progressive wave will advance a distance $vdt$ in time $dt$, during which an amount of charge $dq=CvVdt$ will appear across that portion of the line. Hence the current $I=CvV$ and $Z_0=V/I=1/Cv$. A rigorous proof may be found in Smythe.[1]

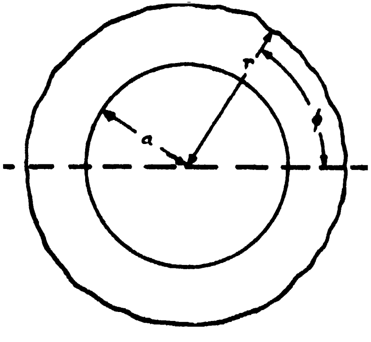


FIG. 1.

While the solution for concentric circles is quite easily obtained, only approximately correct solutions are obtainable for non-circular cables. In this paper we shall consider lines wherein one conductor is circular and the other slightly non-circular. The potential distribution is then calculated as a perturbation of that due to concentric circles. Finally we shall apply these results to the case where one conductor is square.

## CALCULATION OF $C$ AND $Z_0$

In Fig. 1 the inner conductor is a circle of radius $a$ and the outer conductor is specified by its polar equation, $r=f(\phi)$, which can be sufficiently approximated by a Fourier series with $p$ harmonics. Thus,

$$r=f(\phi)=A_0\left(1+\sum_1^p A_j\cos j\phi\right) \qquad (3)$$

with

$$\sum_2^p A_j^2 \ll 1. \qquad (4)$$

The inclusion of sine terms in the Fourier series will not add anything of general value to the discussion. Moreover, most conductors in practice are symmetrical about an axis. Hence we seek a potential distribution $V(r\ \phi)$ such that,

$$\nabla^2 V=0 \qquad (5)$$

with boundary conditions,

$$V=0 \text{ on } r=a \qquad (6)$$

$$=V_1 \text{ on curve (3).} \qquad (7)$$

---

*Consulting Engineering Mathematician.

[1] W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company, Inc., New York, 1939), p. 471.

Let

$$V=L\log r+M+\sum(E_nr^n+F_nr^{-n})\cos n\phi. \tag{8}$$

This expansion[2] will automatically satisfy (5), and to satisfy (6)

$$0=L\log a+M \tag{9}$$

$$0=E_na^n+F_na^{-n}\quad n=1,2,\cdots \tag{10}$$

To satisfy (7), we have after expanding $\log r$, $r^n$, and $r^{-n}$ in Taylor series about $r=A_0$,

$$\begin{aligned}V_1=&L\log A_0+M+L\\&\times[\Sigma A_j\cos j\phi-\tfrac{1}{2}(\Sigma A_j\cos j\phi)^2+\cdots]\\&+\Sigma E_mA_0^m[1+m\Sigma A_j\cos j\phi\\&+\tfrac{1}{2}m(m-1)(\Sigma A_j\cos j\phi)^2+\cdots]\cos m\phi\\&+\Sigma F_mA_0^{-m}[1-m\Sigma A_j\cos j\phi\\&+\tfrac{1}{2}m(m+1)(\Sigma A_j\cos j\phi)^2+\cdots]\cos m\phi,\end{aligned} \tag{11}$$

where . . . denotes higher powers of $A_j$. Now equating coefficients of $\cos n\phi$ in (11),

$$\begin{aligned}-LA_n=&E_nA_0^n+F_nA_0^{-n}\\&+\tfrac{1}{2}\Sigma\Sigma m(E_mA_0^m-F_mA_0^{-m})A_j+\cdots\\&n=1,2,\cdots\end{aligned} \tag{12}$$

and the double summation is taken over $m+j=n$, $m-j=\pm n$.
For $n=0$,

$$\begin{aligned}V_1=&L(\log A_0-\tfrac{1}{4}\Sigma A_j^2+\cdots)+M\\&+\tfrac{1}{2}\Sigma j(E_jA_0^j-F_jA_0^{-j})A_j+\cdots\end{aligned} \tag{13}$$

We now have the linear set of Eqs. (9), (10), (12), and (13) to solve for $L$, $M$, $E_n$, and $F_n$. We should, however, point out that $C$ depends upon $L$ only, since by applying (1) to the inner circle, the line integral of the cosine terms in (8) vanish; thus

$$C=2\pi KL/V_1. \tag{14}$$

We now solve Eqs. (10) and (12) for $E_n$ and $F_n$ by successive approximations. For a first approximation, we neglect the double summation and higher order terms in (12) and solving, we have,

$$E_n=-\frac{A_nL}{a^n[(A_0/a)^n-(a/A_0)^n]}, \tag{15}$$

$$F_n=-\frac{a^nA_nL}{(A_0/a)^n-(a/A_0)^n}. \tag{16}$$

Substituting these expressions in (13) and subtracting (9) from (13), we have to a first approximation,

$$C=\frac{2\pi K}{\log A_0/a-\sum_1 A_j^2(\frac{1}{4}+\frac{1}{2}jR_j)}\quad\text{farads/meter} \tag{17}$$

$$R_j=\frac{(A_0/a)^j+(a/A_0)^j}{(A_0/a)^j-(a/A_0)^j}. \tag{18}$$

Thus we have an approximation which is of the order of $jA_j^2$. We can now obtain further approximations by substituting (15) and (16) in the neglected part of (12) and resolving for $E_n$ and $F_n$, and so forth. This, however, will result in corrections to $C$ which are of higher powers in $jA_j^2$ and hence may be neglected because of (4). It should also be noted that $A_0$ is always greater than $a$; otherwise there would be points on the outer conductor for which $r$ is less than $a$, which is impossible.
Also, by (2),

$$Z_0\text{ in air}=60[\log A_0/a-\sum_1 A_j^2(\tfrac{1}{4}+\tfrac{1}{2}jR_j)]\text{ ohms} \tag{19}$$

since $v=3\times10^8$ meters/sec., and $K=8.85\times10^{-12}$ farads/meter.

A case of interest is where $A_0/a$ is of the order of 3 or greater. Then $R_j$ is approximately 1 and

$$Z_0=60[\log A_0/a+d] \tag{20}$$

$$d=-\sum_1 A_j^2(\tfrac{1}{4}+\tfrac{1}{2}j). \tag{21}$$

Since $d$ is at most of the order of a few percent, we may replace $d$ by

$$\log(1+d)\text{ and }Z_0=60\log A^*/a \tag{22}$$

and $A^*$, which is $A_0/1+d$, may be considered as the outer radius of an equivalent coaxial (concentric lines) line, with the same characteristic impedance.

### APPLICATION OF RESULTS TO SQUARE OUTER CONDUCTOR

For a square of side $2b$, the polar equation is,

$$\begin{aligned}&r=b\sec\phi\quad 0\le\phi\le\pi/4\\&r=b\csc\phi;\ \pi/4\le\phi\le\pi/2\end{aligned} \tag{23}$$

---

[2] See reference 1, p. 63.

with a periodicity of $\pi/2$. Hence only $\cos 4j\phi$ terms will appear in the Fourier series, and

$$A_0=\frac{4b}{\pi}p_0$$
$$A_0A_{4j}=\frac{8b}{\pi}p_j;\quad p_j=\int_0^{\pi/4}\sec\phi\,\cos 4j\phi d\phi. \tag{24}$$

Now,

$$p_{j-1}-p_j=2\int_0^{\pi/4}[\cos(4j-3)\phi-\cos(4j-1)\phi]d\phi$$
$$=(-1)^{j+1}2.828/(4j-3)(4j-1).$$

Thus knowing $p_0$, $p_j$ is easily calculated; and $p_0=.881$, $p_1=-.062$, $p_2=.019$, $p_3=-.010$, $p_4=.005$, $A_0=1.12b$; $A_4=-.14$, $A_8=.044$, $A_{12}=-.021$, $A_{16}=.012$. Thus condition (4) is approximately satisfied with $p$ equal to 12. Thus,

$$Z_0=60[\log b/a+.108 - 039R_4-.008R_8-.002R_{12}], \tag{25}$$

with

$$R_4=\frac{1.570(b/a)^4+.636(a/b)^4}{1.570(b/a)^4-.636(a/b)^4}, \tag{26}$$

and $R_8$ and $R_{12}$ are given by similar expressions, with the constants replaced by (2.480, .403) and (3.870, .258), and the exponents by 8 and 12, respectively.

For $b/a$ of the order of 2, the equivalent outer radius defined by (22) is $1.06b$. This value compares more or less favorably with $1.08b$, obtained by Frankel.[3] For $b/a$ less than 2, Frankel's derivation does not apply, and the correct $Z_0$ is given by (25).

For $b/a$ equal to 1, $Z_0$ from (25) is .06 ohms, which is fairly close to the actual value of $Z_0$, namely, 0. This constitutes a partial check upon the accuracy of this formula.

Finally I wish to thank Mr. Benjamin Parzen for suggesting the problem.

---

[3] S. Frankel, "Characteristic Impedance of Parallel Wires in Rectangular Troughs," Proc. I.R.E. **30**, 182 (1942).

---

# Thermal Conductivity of Aluminum; Solid and Liquid States

C. C. Bidwell and C. L. Hogan
*Lehigh University, Bethlehem, Pennsylvania*

(Received April 9, 1947)

Thermal conductivity measurements on lead, tin, and zinc extending over the liquid as well as the solid states were reported by one of the authors in 1940. The present report extends this work to cover aluminum in both the liquid and solid phases. An improved technique is described. The law $k/\rho C=K(1/T)+K'$ is found to hold for aluminum also, with the intercept, $K'$, the same for both states. The thermal conductivity ($k$) decreases with rise of temperature as it does for lead, tin, and zinc. On this point previous workers have published very conflicting data.

## INTRODUCTION

Data on the thermal conductivity of aluminum in the handbooks, the critical tables, and as collected and published in *The Aluminum Industry*[1] are conflicting particularly as to the change with temperature. No data for the liquid state are reported. Several observers report a rise of thermal conductivity with temperature; only one, Konno,[2] reports a decrease. The present data on a specimen of purity 99.20 percent is in exact agreement with Konno's values (0.50 at 0°C; 0.36 at 600°C). A specimen of purity 99.95 percent is found to have a higher conductivity, *viz.*, 0.54 at 25°C, decreasing to 0.44 at 657°C. On this specimen, measurements were extended into the liquid state at points 740°C and 900°C.

## THE METHOD

The Forbes bar method previously used was further modified and adapted to high temperature work. A cylinder of the material under study, 25 cm long, 2.5 cm in diameter, was

---

[1] Edwards, Frary, and Jeffries, *The Aluminum Industry* (McGraw-Hill Book Company, New York, 1930).
[2] S. Konno, Phil. Mag. **40**, 542 (1920).

centered vertically in a tubular furnace and surrounded with finely screened silocel. Thermojunctions were previously fastened to the rod every 3 cm by insertion into small holes drilled about 1.5 mm deep, and the junctions peened in. At the upper end, just above the top junction, was placed a heater coil of about 10 turns of B & S No. 26 nichrome wire, insulated from the bar by mica or asbestos. The junction wires, encased in fine quartz tubes, were led down and out the bottom of the furnace tube, the heater wires going out the top. A reference junction was inserted into the furnace from below to a point just below the bottom of the test rod but not touching it. Another reference junction, as a spare, was inserted from above to a point about half-way down the specimen but placed near the furnace tube so as not to be affected by the heater at the top of the rod. The junctions could be connected in turn differentially with either of the reference junctions and led directly to the galvanometer (Leeds and Northrup Type R 2500 a). At any particular temperature steady conditions were indicated by the galvanometer readings. (The zero was adjusted to the right hand end of the scale at 30, thus making the whole scale available.) Deflections were differentials between the given junction and the reference. At room temperatures all junctions would read alike at 30. At higher temperatures there was always a residual temperature difference, and initial readings for each junction had to be taken and the heater turned on only when steady readings were obtained. A deflection of 60 divisions meant approximately 6°C. At upper temperatures the deflection for the bottom junction was very small, while that for the top junction (No. 1) might be as much as 20 divisions. The elimination of this gradient while desirable was not necessary. Also a slight drift of all junction readings due to temperature drift had to be tolerated and corrected for.

## PROCEDURE

When all junctions show fairly steady condition, the time is noted, and the readings from top to bottom of the specimen and back again are recorded and again the time noted. This is repeated several times in order to determine the drift. Then the time is again noted, the heater turned on and, after about three minutes, the readings again taken from top to bottom and back and the time noted before and after. This is immediately repeated, and then the switch is opened. The cooling data are immediately obtained, using the top junction and taking readings every half minute. The readings should all be taken within about ten minutes from the time of closing the heater switch, otherwise the reference points become uncertain. Only the cooling data from the top junction are significant, the other junctions indicating too low a rate because of the inflow of heat from the upper part of the bar. The cooling rate for the top junction (which is about 4 cm from the top of rod) is approximately corrected with respect to the heat flow down the rod although for the first minute even this rate is low. This shows up on the $dT/dt$, $T$ curve which is a straight line except possibly for the first few points. We have now two sets of data, the gradient data and the cooling data. The actual temperature is observed by connecting one of the reference junctions (or any junction) to a potentiometer. The gradient data and the cooling data are now plotted. First the temperatures (galvanometer deflections) are plotted for the gradient data, the deflections as ordinates, the distances down the rod as abscissa, taking the position of the top junction as $x=0$. This is the $T$, $x$ curve. Slopes of this curve are measured graphically and the $dT/dx$, $x$ curve plotted. Then the slopes of this curve are measured and the $d^2T/dx^2$, $x$ curve plotted. The cooling data are plotted giving the $T$, $t$ curve ($T$ as ordinates and $t$ (time) as abscissa). The slopes of this curve are measured and plotted, $dT/dt$ as abscissa, $T$ as ordinates. Corresponding* values of $dT/dt$ and $d^2T/dx^2$ for the various values of $T$ are plotted, $d^2T/dx^2$ as ordinates, $dT/dt$ as abscissa. The straight lines usually go through the origin but not always. The constant value of $dT/dt/d^2T/dx^2$ is called $m$ (actually the reciprocal of the slope considering the way the curves are plotted), and $m=k/\sigma\rho$, where $k$ is the thermal conductivity, $\sigma$ the specific heat, $\rho$ the density. $k$ is thus determined if $\sigma$ and $\rho$ are

---

* Corresponding values are values for the same temperature $T$. $dT/dt$ values are read directly, but $d^2T/dx^2$ is found by finding the value of $x$ for the given $T$ (on the $T$, $x$ curve) and for this value of $x$ moving vertically to the value of $d^2T/dx^2$.

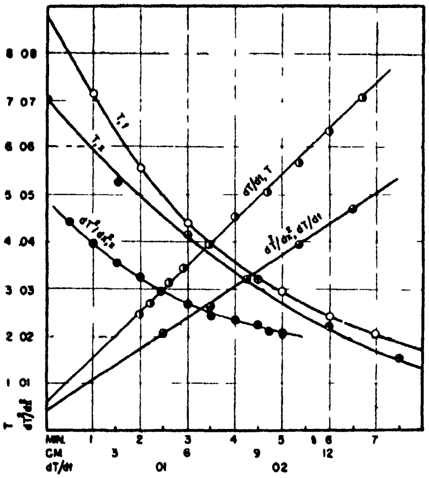


FIG. 1. Sample set of data for thermal conductivity of aluminum at 650°C.

known. Figure 1 is a sample of the plots. A similar set of curves is obtained for each value of $k$ determined.

## THEORY

This is the old Forbes method described by Forbes in 1868. The difference in the heat flowing down the rod at one level and out at a slightly lower level is $kA(dT/dx)_1 - kA(dT/dx)_2$, and this difference is $\sigma\rho Al dT/dt$. $A$, the cross section cancels, $l$ is the difference between the two levels, and

$$\frac{(dT/dx)_1 - (dT/dx)_2}{l} = d^2T/dx^2.$$

The equation is

$$\sigma\rho dT/dt = kd^2T/dx^2$$

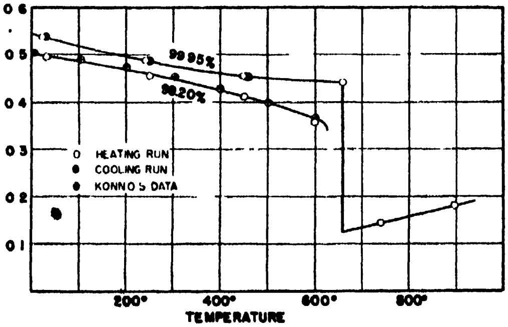


FIG. 2. Thermal conductivity and temperature of aluminum.

or

$$dT/dt = \frac{k}{\sigma\rho} d^2T/dx^2.$$

This calls for a straight line through the origin. Often the curve does not go through the origin. In taking the data one has to compromise—one cannot wait until conditions are completely steady. The reference junction readings are slowly changing, and hence one must be satisfied even though the temperatures are still slightly rising. This introduces a term $\sigma\rho dT/dt_H$ on the left side of the above leading to the equation

$$\frac{dT}{dt_C} = \frac{k}{\sigma\rho}\frac{d^2T}{dx^2} - \frac{dT}{dt_H}.$$

$\sigma\rho dT/dt_H$ is the residual heating, $\sigma\rho dT/dt_C$ is the observed cooling. Careful study of the data shows that the intercept is thus accounted for.

When measurements in the liquid state are desired the specimen must be contained in some inert container. In this work graphite was used—a long cylindrical crucible being bored out of a graphite electrode. The walls were about 3 mm thick, and the junctions were placed in small holes drilled into these walls but not clear through. The wires encased in quartz tubes were tied on with threads, the silocel packing holding them in place when the threads burned away. The heater was wound on the graphite cylinder (insulated with mica) and measurements made as before. The equation now is

$$\sigma_a\rho_a A_a dT/dt + \sigma_g\rho_g A_g dT/dt = k_a A_a d^2T/dx^2 + k_g A_g d^2T/dx^2$$

or

$$\frac{dT}{dt} = \left[\frac{k_a A_a + k_g A_g}{\sigma_a\rho_a A_a + \sigma_g\rho_g A_g}\right]\frac{d^2T}{dx^2}.$$

If one takes account of the slight heating, the term $dT/dt_H$ is to be subtracted from the right side as above. The subscripts refer to aluminum and graphite. Values of $\sigma_a\rho_a$ were obtained from *The Aluminum Industry* or determined experimentally. The latter was necessary for the graphite. For this purpose observations were made as above on a graphite rod at room temperature and at 700°C. This gave values of $k_g/\sigma_g\rho_g$. $\sigma_g\rho_g$ was obtained by an auxiliary experiment in which a graphite rod with an axial heater was held at various temperatures and the watts input noted. Later cooling curves were taken through these temperatures. The equation $\sigma_g\rho_g A l dT/dt = VI/4.18$ permits the computation of $\sigma_g\rho_g$.

## RESULTS

The values of the thermal conductivity of aluminum (99.95 percent) are as follows: at 25°C, 0.538; at 240°C, 0.486; at 250°C, 0.486; at 450°C, 0.452, 0.456; at 650°C, 0.445; at 740°C (liquid), 0.143; at 900°C (liquid), 0.180. For the aluminum (99.2 percent**) the values are: at 25°C, 0.500; at 245°C, 0.457; at 450°C, 0.414; at 590°C, 0.365. These values are shown graphically in Fig. 2, together with Konno's points. The error in these values does not exceed ±.005, assuming the accuracy of the values of $\sigma\rho$ (taken from *The Aluminum Industry*). The 99.95 specimen shows higher conductivity throughout, and the change with temperature is a decrease paralleling the change for the commercial specimen. Figure 3 shows the data for the c.p. specimen plotted to test the relation $k/\rho C = K(1/T) + K'$. This relation is found to hold with the intercept, $K'$, the same for the solid and liquid phase, exactly as was found for tin, lead, and zinc.[3] ($C$ in this equation is the atomic heat.) Since only two points were found for the liquid state it probably is more proper to state that the data are at least consistent with this equation.

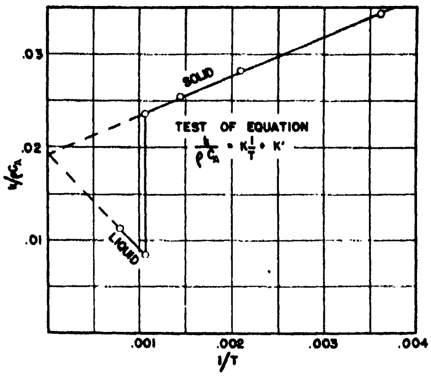


FIG. 3.

## THERMAL CONDUCTIVITY OF GRAPHITE

The thermal conductivity of graphite as here determined was found as 0.30 at 25°C. Buerschaper[4] finds 0.26 for radial and 0.40 for longitudinal, while Powell and Schofield[5] find 0.34 at this temperature. The value here found for 700°C, 0.123, is consistent with an extrapolation of the data of the above observers.

## ACKNOWLEDGMENT

The aluminum specimens and the analyses were furnished by The Aluminum Company of America together with a grant of funds to facilitate this work. For this help the authors express their appreciation to this Company.

---

** Analysis of the 99.2 percent specimen: Si 0.10 percent, Fe 0.67 percent, Cu 0.01 percent, Mn<0.01 percent, Mg<0.01 percent.

[3] Bidwell, Phys. Rev. **58**, 561 (1940).

[4] Buerschaper, J. App. Phys. **15**, 452–454 (1944).

[5] Powell and Schofield, Proc. Phys. Soc. **51**, 153–172 (1939).

# Letters to the Editor

## Stabilization, Tempering, and Relaxation in the Austenite-Martensite Transformation *

J. H. HOLLOMON,† L. D. JAFFE, AND D. C. BUFFUM
*Watertown Arsenal Laboratory, Watertown, Massachusetts*

Received May 12, 1947

TRANSFORMATION in steels of the face-centered cubic phase, austenite, to the body-centered tetragonal phase, martensite, ordinarily takes place only during cooling. The temperature (*Ms*) at which the transformation begins is independent of the cooling rate. The extent of transformation at any temperature below *Ms* is less, the lower the cooling rate.[1] Moreover, holding at a constant temperature below *Ms* decreases the extent of transformation on subsequent cooling to a fixed lower temperature.[2] This phenomenon, termed stabilization, is believed to arise from a change in microstresses with time. This change might result from (*a*) decomposition of the metastable martensite (tempering) and consequent dimensional change, or (*b*) stress relaxation (viscous flow) of microscopic regions.

Specimens of a quenched steel containing 1 percent C, 1.5 percent Cr, were stabilized at various temperatures and times. The specific volume at room temperature (*R.T.*) was then determined before and after cooling to −196°C. The results (Fig. 1) indicate that the time and temperature for a fixed degree of stabilization are related through an activation energy of about 21,500 cal./mole.

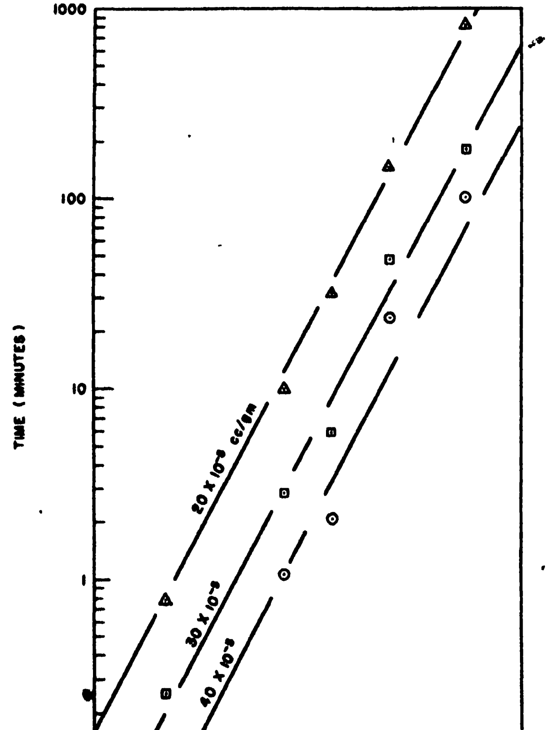


FIG. 1. Relation between time and temperature of stabilization for a fixed change of specific volume upon subatmospheric cooling.

Specimens of a quenched plain-carbon, 0.95 percent C, steel were cooled to −196°C, then tempered at various temperatures and times. The tetragonality, $c/a-1$, of the martensite lattice at *R.T.* was determined by x-ray diffraction methods. The results (Fig. 2) indicate an activation energy of 34,000 cal./mole for the first stage of tempering.

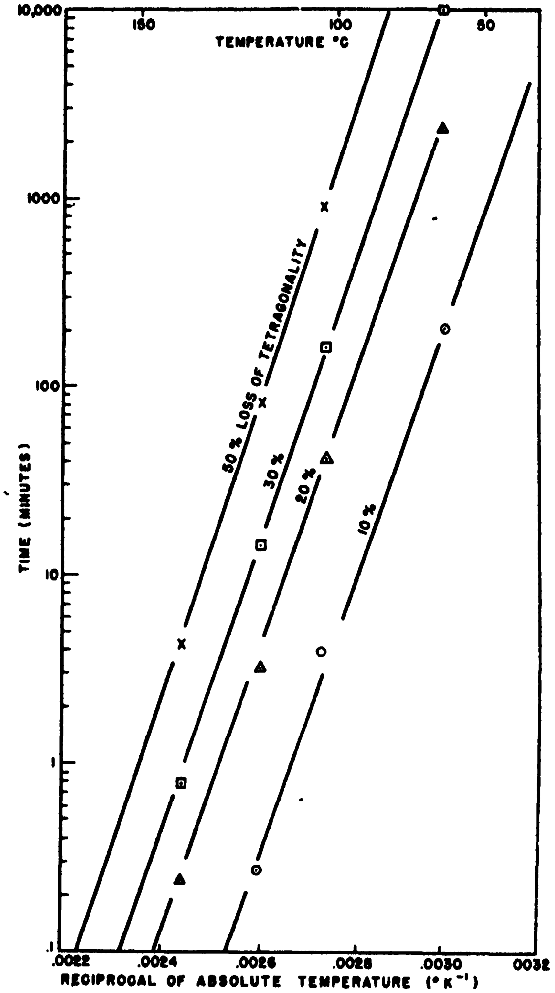


FIG. 2. Relation between time and temperature for fixed loss of tetragonality.

Rods of quenched plain-carbon steel, 1.25 percent C, were heated to 100°C and twisted a fixed amount, not enough to cause permanent set. After holding 30 min., the specimens were cooled to lower temperatures and released. The anelastic twisting, measured as a function of time

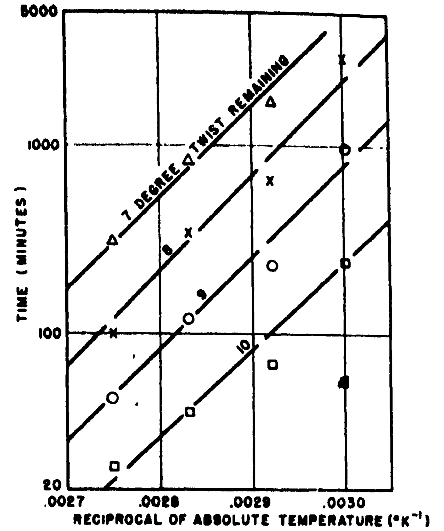


FIG. 3. Relation between time and temperature of untwisting for a fixed amount of untwisting. Twisted 30 degrees at 100°C. Held 30 minutes.

(Fig. 3), corresponds to an energy of about 22,500 cal./mole for stress relaxation. The energies observed for stabilization and relaxation are approximately equal to that reported [3] for stress relaxation in plastically deformed iron (ferrite).

Different steels were used for the different tests to obtain the required accuracy. However, changes in steel composition are unlikely to affect the activation energies materially, particularly in the light of the above agreement between the heats of activation for stress relaxation. It is concluded that stabilization in the austenite-martensite reaction does not arise from tempering of martensite, and probably arises from stress relaxation. Apparently, the relaxation centers responsible for stabilization are produced by local plastic flow caused by austenite-martensite transformation and perhaps by non-uniform thermal contraction.

This work will be discussed in detail elsewhere.

---

[*] The statements and opinions in this letter are those of the authors and do not necessarily express the views of the Ordnance Department.

[†] Now at General Electric Research Laboratories, Schenectady, New York.

[1] J. A. Matthews, Trans. A.I.M.E. 71, 568 (1925).

[2] G. Tammann and E. Scheil, Zeits f. anorg. allgem. Chemie 157, 1 (1926).

[3] W. A. West, Tech. Pub. 1993, A.I.M.E., Metals Tech. 13, No. 5 (1946).

---

## Remarks on Compressive and Tensile Strengths

Concerning a Paper by P. W. Bridgman

STANLEY THOMPSON
*North American Aviation, Inc., Los Angeles, California*

April 9, 1947

IN a recent paper[1] Bridgman has discussed a series of tests on the effect of hydrostatic pressure on the fracture of brittle materials. The remarks of this letter are considered pertinent to Bridgman's paper, since they explain some of the anomalies mentioned.

Tests to failure in compression by Föppl[2] indicate that the magnitude of the compressive strength, as well as the type of failure observed, depends markedly on the end conditions of loading. Föppl tested his specimens under two end conditions of loading, lubricated and unlubricated loading faces; and with two types of compressive loading, one-dimensional and two-dimensional. With a specimen of concrete, the compressive strengths in the four cases were related as shown in Fig. 1.

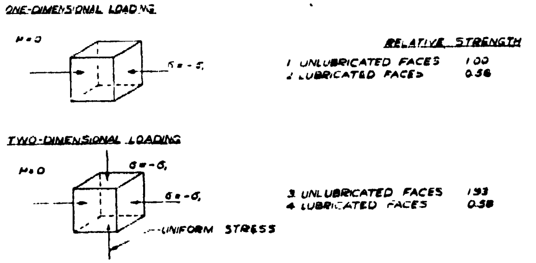


FIG. 1.

With the surfaces unlubricated, failure occurred by breaking off wedge-shaped pieces, while with lubricated surfaces failure occurred in planes parallel to the direction of compression—as if a tension existed at right angles to the direction of compression—with almost the same load for the one-dimensional case as for the two dimensional. The variation between lubricated and unlubricated surfaces is attributed by Föppl to the effect of shear stress across the loading faces. Since a compressive specimen must be comparatively short to prevent buckling, the end conditions are rarely negligible.

For an analysis of the type of failure produced and a prediction of the comparison between compressive and tensile strengths, it is desirable to consider the load systems shown in Fig. 2. Figure 2(a) shows a two-dimensional compressive loading $(-\sigma_1)$; Fig. 2(b) shows a hydrostatic pressure, numerically equal to the compressive loading of Fig. 2(a) on which is superimposed a one-dimensional tensile loading numerically equal to the hydrostatic pressure. Neither the theory of elasticity, nor any test, can differentiate between these two loading systems, since they are physically identical.

In that which follows, aside from the entire equivalence of load systems (a) and (b) in Fig. 2, one assumption is made: the equivalence of breaking stress under one- and two-dimensional loading, as found experimentally by Föppl for lubricated loading faces, and as predicted by the

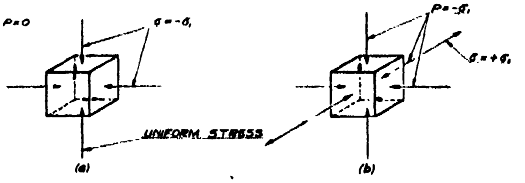


FIG. 2.

semi-empirical shear and octahedral-stress theories. As a result any given one-dimensional compressive breaking stress ($-\sigma_1$) superimposed on a given hydrostatic pressure ($-P_1$) is equivalent to a one-dimensional tensile breaking stress ($+\sigma_1$) of equal magnitude superimposed on a hydrostatic pressure ($-P_1-\sigma_1$). A plot satisfying this condition of breaking stress $\sigma(P)$ against hydrostatic stress $P$ is shown schematically in Fig. 3. Positions above the horizontal axis represent tensile breaking stresses, those below compressive breaking stresses; positions to the right of the vertical axis represent hydrostatic tensions (triaxial tension), those to the left, hydrostatic pressure (triaxial compression). Tensions are considered positive, compressions negative. It should be noted that $\sigma_P$ is zero when $P$ is positive, which corresponds to failure under hydrostatic (triaxial) tension only.

The relationship shown in Fig. 3 can be expressed mathematically by the equation

$$f(P) = -g(P+f), \tag{1}$$

where $f(P)$ represents either branch of the curve for breaking stress $\sigma_P$, and $g(P+f)$ represents the other branch.

It is interesting to compare the slope $df(P)/dP$ of the branch $f(P)$ at any abscissa $P$ with the slope $dg(P+f)/d(P+f)$ of the branch $g(P+f)$ at the corresponding $(P+f)$. From Eq. (1):

$$\frac{dg(P+f)}{d(P+f)} = -\frac{\dfrac{df(P)}{dP}}{1+\dfrac{df(P)}{dP}}. \tag{2}$$

These slopes correspond to the pressure coefficients of stress for the two branches of the curve. It can be concluded from Eq. (2) that a pressure coefficient of tensile strength at $P$ between 0 and 1 corresponds to a pressure coefficient of compressive strength at $(P+f)$ between 0 and $\infty$, and the coefficient for compressive strength is always the larger.

Bridgman reports that "the effect of hydrostatic pressure is markedly greater in increasing the compressive strength of glass than in increasing its tensile strength." Equation (2) and Fig. 3 give a measure of this effect for any function $f(P)$. If either the compressive or tensile breaking stress is known experimentally as a function of hydrostatic stress, the other may be obtained without experiment by use of Eq. (1) or Fig. 3.

Since actual compressive failures of homogeneous materials under hydrostatic pressure cannot occur, it is concluded that failure of a material under compressive loading, in one or two directions, will occur in a manner typical of tensile failures for that material under corresponding hydrostatic pressure. Any departure from this type of failure, or from behavior predicted by a plot similar to Fig. 3, is a measure of the experimental departure from idealized end conditions or inhomogeneity of material. It is believed that the angular fracture of glass in simple compression when supported by hydrostatic pressure, shown in Bridgman's Fig. 3, may be due to the end effects investigated by Föppl and that the numerical results thus obtained for compressive strength are unreliable.

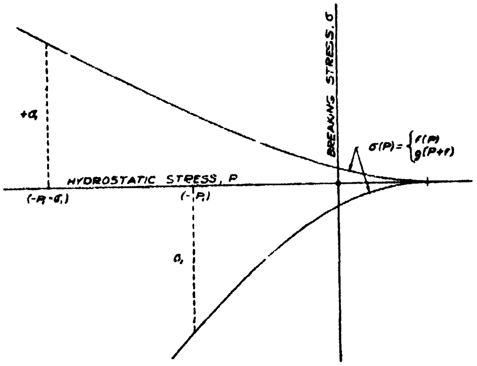


FIG. 3.

---

[1] P. W. Bridgman, J. App. Phys. 18, 246 (1947).

[2] A. Foppl, Mitteilungen ausdem Mech. Tech. Lab. in München (1900).

---

# Reply to "Remarks on Compressive and Tensile Strengths"

P. W. Bridgman
*Harvard University, Cambridge, Massachusetts*

April 22, 1947

THANK you for letting me see the remarks by Stanley Thompson on my paper on brittle fracture under pressure. These remarks are interesting as showing the consequences of carrying through to a mathematical conclusion the implications of a literal acceptance of the results of the equivalence of one and two dimensional loading in compression. I think, however, that one would not expect the bearing on the experimental situation to be any too close. Föppl's results were obtained only for a single material of not very typical properties, namely, cement and furthermore were obtained only at atmospheric pressure. In developing his mathematics Thompson had to make an additional assumption not covered by the experiments of Föppl, namely, that the equivalence of one and two dimensional compression would continue to hold if the experiment were performed in a medium to which any hydrostatic pressure whatever is applied in addition to the compressions. I very much question whether this is legitimate. One of the conclusions that could be drawn is that when the effects of terminal friction are eliminated, only one type of fracture occurs, the same under simple tension as under simple compression. It is tacitly assumed that this type of fracture is the simple tension type, but I can see no intrinsic reason in the argument for preferring this to the ordinary simple compression type except the highly special experiments of Föppl. The point of view of Thompson also commits him to the position that the same type of fracture will prevail along the entire course of both the upper and lower branches of the curves in his Fig. 3. This seems to me highly improbable.

Of course I will have to admit that there is some justice to his last sentence, namely, that terminal friction modifies the figures for the fracture stresses that would be given by using only the figures for hydrostatic pressure and compressive stress. But, I think the modification would not be large, and I think the effect is in the wrong direction. The most complete fracture experimentally was obtained near the ends, where the effect of friction in fictitiously raising the strength is the greatest according to Thompson's suggestion. If this is correct it seems to me that fracture should have occurred first at the center and should have been most complete there. Furthermore, I think substitution of the numerical values into a diagram like his Fig. 3, both for glass and Carboloy, would show that a rather improbable state of affairs is demanded.

Certain aspects of Thompson's analysis of the stress system into hydrostatic pressure and superposed simple stresses are much like my analysis in Mechanical Engineering, February 1939, in which similar questions are discussed.

## European Induction Accelerators

Rolf Wideröe
*Zürich, Switzerland*

(April 21, 1947)

I have read with interest in the January number of this journal H. F. Kaiser's account concerning "European Induction Accelerators" and should be grateful if you would allow me to augment and also to correct some of the statements contained therein.

First of all, as regards the historical data concerning the development of the betatron: in the autumn of 1922 I had already invented the "ray transformer" (my own designation for the electron induction accelerator) in Karlsruhe during my studies there, quite independently of Slepian. In 1925 I applied for a German Patent on this invention, but this patent application was subsequently abandoned. After the completion of my experimental work under Professor Rogowsky in Aachen (1925–1927),[1] it was only in the autumn of 1942 that I learned of the work done by Kerst and other investigators in the meantime.[2] Subsequently, up to June 1943 I was able to make a number[3] of inventions during my work at the Norsk Elektrisk Brown Boveri in Oslo, these inventions being confiscated in the autumn of 1943, however, by the German authorities. Slepian's U. S. Patent, which had not been mentioned by Kerst in his first publication, only came to my knowledge accidentally some time later, so that it appears to me that Slepian's proposals can only have been of slight importance as regards the development of the induction accelerator.[4]

In Kaiser's account it is stated that the formula for the average electron ray current was found by me. This is an error, and in my work[5] it is clearly stated that this formula and its derivation originate from Kerst. According to this formula it should not be possible to obtain an appreciable amount of $\gamma$-rays with a 15-mv transformer at such a low frequency as 50 cycles. With $\Delta r=2$ cm and $dr=0.1$ cm, for instance, the formula gives an injection voltage of about 170 v and an average current strength of only about 0.0085 $\mu$A. Compared with this we used an optimum injection voltage of 7–10 kv and attained an average current of about 0.03 $\mu$A, the maximum obtainable current strength for this injection voltage being probably about 10 times higher. The remarkable fact that small ray transformers can still be operated to advantage with such a low frequency as 50 cycles was entirely novel in 1944 and certainly represents the most important result obtained from the work performed in Hamburg.

Our experimental results obtained with the 15-mv transformer can, as shown by Touschek and myself, be explained by elaborating the original Kerst-Serber assumptions.[6] When in accordance with our theory the electron source is placed inside the potential channel of the stabilizing forces, the efficiency of the electron injection can be calculated approximately from:

$$\eta=\mathrm{Eksp}\left(\frac{-16\sqrt{2}}{3\pi^2}\,\frac{Ui}{\Delta U}\left(\frac{b}{2\Delta r}\right)^{\frac{1}{2}}\arcsin\frac{h}{2\Delta z}\right),$$

where

$Ui$ = injection voltage
$\Delta U$ = induced voltage per turn
$b$ = width of electron source
$h$ = height of electron source
$\Delta r$ = radial oscillation amplitude of electrons
$\Delta z$ = axial oscillation amplitude of electrons.

If the electron gun is located outside the potential channel ("injection from outside"), the efficiency of the electron injection will depend on the divergency of the injected electron beam, because only very small deviations from the optimum direction of injection are permissible for catching the electrons. With the 15-mv transformer, the angular deviation for $Ui=10$ kv may for instance only amount to about 0.16°–0.32°. It can be shown that in this case the efficiency of the injection in first approximation will be proportional to:

$$\left(\frac{\Delta U}{U_i}\right)^x \text{ where } \tfrac{1}{3}\leqq x\leqq\tfrac{3}{4}.$$

In conclusion I should like to state that our work in Hamburg was performed quite independently and without knowledge of that done by Siemens-Reiniger and other investigators. An account given at a mutual discussion in Erlangen (Nov. 1944) concerning the favorable results obtained by us with the 50-cycle ray transformer probably may have given the impetus for the later construction of corresponding 50-cycle apparatus by Siemens.

---

[1] Described in Achiv für Elektrotechnik **21**, 387 (1928).

[2] Phys. Rev. **59**, 110 (1941); **60**, 47 (1941).

[3] E.g., "Premagnetising the control field," German Pat. Appl. W 113740 VIIIc/21g. "Elimination of the air gap by reverse magnetisation," German Pat. Appl. W 113763 VIIIc/21g. "New injection system for electrons," German Pat. Appl. W 113563 VIIIc/21g. "Electron lenses," German Pat. Appls. W 113742 and W 113925 VIIIc/21g.

[4] I presume that the well-known phenomenon of an electrodeless annular discharge, developing in a poor vacuum chamber under the influence of high frequency magnetic fields, provided the first impulse for my own inventions.

[5] Archiv für Elektrotechnik **37**, 391–408 (1943).

[6] A detailed report of this work will shortly appear in the "Schweizer Archiv."

## Direction of Maximum Crystal Elongation during Metal Cutting*

G. H. TOWNEND
*Metropolitan-Vickers Electrical Company, Ltd., Manchester, England*
March 17, 1947

IN an interesting paper by Dr. M. E. Merchant on "Mechanics of the Metal Cutting Process,"[1] an error was made in the expression connecting shear angle, rake angle, and direction of maximum crystal elongation (Eq. (1), page 269). I have written to Dr. Merchant and he confirms the new expression derived below.

Let $\alpha$ = rake angle,

$\phi$ = the angle between the shear plane and the surface of the workpiece.

Consider a section of the workpiece $ABCD$ such that $< CBD = 90°$. Then after shearing this section will become the parallelogram $BFEC$. The dimension perpendicular to the shear plane must be unchanged during shearing.

$$\therefore\quad BD = BH = a.$$

Consider a line $CG$ in the chip making an angle $\psi$ with the plane of shear. Then the corresponding line in the workpiece is $DJ$. The condition that $CG$ shall be the direction of maximum crystal elongation is that $CG/DJ$ is a maximum.

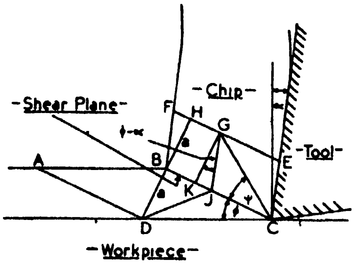


FIG. 1.

$$CG = a\ \mathrm{cosec}\psi,$$
$$DJ = \{a^2 + (a\cot\phi - a\cot\psi + a\tan(\phi-\alpha)^2)^2\}^{\frac{1}{2}}.$$

The condition is that

$$\{\sin^2\psi + (\cot\phi\sin\psi - \cos\psi + \tan(\phi-\alpha)\sin\psi)^2\}^{\frac{1}{2}}$$

is a maximum. That is,

$$\sin^2\psi + \{\sin\psi(\cot\phi + \tan(\phi-\alpha) - \cos\psi)\}^2$$

is a minimum. Let

$$\cot\phi + \tan(\phi-\alpha) = \epsilon,$$
$$\epsilon^2\sin 2\psi - 2\epsilon\cos 2\psi = 0.$$

If $\epsilon \neq 0$

$$2\cot 2\psi = \epsilon,$$
$$2\cot 2\psi = \cot\phi + \tan(\phi-\alpha).$$

Alternatively it is easily shown that a circle on the workpiece becomes an ellipse on the chip and that the principal axes of the ellipse make an angle $\psi$ with the shear plane, where

$$2\cot 2\psi = \cot\phi + \tan(\phi-\alpha).$$

---

* This letter was published in the May issue of the Journal of Applied Physics, but through an error Fig. 1 was omitted.

[1] J. App. Phys. 16, 267 (1945).

# Here and There

### New Appointments

A recent issue of *Science* reported the following new affiliations:

E. C. Watson, Professor of Physics at California Institute of Technology, has been appointed chairman of the Division of Physics, Mathematics, and Electrical Engineering.

C. J. Gorter, Professor of Physics and Co-Director, Physical Institute, University of Leiden, and formerly Zeeman Professor of Physics, University of Amsterdam, is a visiting lecturer at Harvard University for the 1947 summer term. He is conducting one course in advanced dynamics and one in relaxation phenomena. The latter treats magnetic and dielectric absorption and dispersion in alternating fields, a subject in which Professor Gorter has made outstanding contributions.

H. P. Robertson, a member of the faculty of Princeton University since 1928, has been appointed Professor of Mathematical Physics, California Institute of Technology, effective July 1.

### Distribution of Concentrated Boron 10

Concentrated isotope boron 10, which enables scientists to make more sensitive neutron-detection tubes and thereby speed up results in atomic research problems, is now available, the Isotopes Branch of the Atomic Energy Commission has announced. However, only a limited quantity is ready for general distribution. Requests will be carefully reviewed and allocations will be restricted to reasonable quantities for the proposed investigation.

Boron 10 atoms are normally quite stable. They have, however, a great affinity for neutrons. When a boron 10 atom captures a neutron it is transmuted into two energetic nuclear particles, one lithium and the other helium. The latter is the well-known alpha-particle. Either particle serves as an excellent trigger for a tube designed to count neutrons. Boron found in nature contains only 20 percent boron 10. Hence a neutron-detection tube five times more sensitive can be made by employing nearly pure boron 10 rather than an equal quantity of ordinary boron. The boron 10 was produced by concentrating the isotope from natural boron in studies of isotope concentration.

Boron 10 will be packaged for shipment in the form of the solid complex boron trifluoride-calcium fluoride. The complex contains 6.9 percent elemental boron, of which 96 percent is B 10. Approximately 6.5 grams of $BF_3 \cdot CaF_2$ are needed to obtain one liter of $BF_3$ at normal temperature and pressure (assuming 100 percent liberation). The boron trifluoride may easily be released as a gas by heating to temperatures above 260°C in a vacuum.

The price of $BF_3 \cdot CaF_2$ complex is $2 per gram, independent of quantity. Shipping charges will be added to the invoice. There is no additional handling fee per shipment. Standard units of 1, 5, 10, and 50 grams have been packaged in glass containers with moisture-proof plastic screwtops. Requirements will be filled in single units or multiples of these standard sizes. Clinton Laboratories, Oak Ridge, Tennessee, will act as the supplier.

### Exhibit of Technical Photographs

A special section devoted to photographs of technical and scientific subjects will be included again this year in the annual exhibition of the Photographic Society of America. The exhibition will be hung at the Oklahoma Art Center, Oklahoma City, and contributors are invited to submit prints for the technical section prior to September 8. Prints should be mailed to W. F Swann, 343 State Street, Rochester 4, New York

Subject matter may cover any phase of technical photography save pictorial photographs of technical or mechanical operations. Scientific and industrial photographs illustrating original and novel photographic techniques may be submitted. These include photographs in such fields as astronomy, geology, medicine (except diagnostic radiographs), metallography, mineralogy, and physics. Both black-and-white or color photographs are acceptable. Technical data pertaining to each photograph should be lettered neatly in the lower left-hand corner of each mount, if possible, and the maker's name should appear on the face of the mount. There is no limit to the number of prints which may be submitted by a competitor.

### Brookhaven Advisory Committee

Dr. Edward U. Condon, Director of the National Bureau of Standards, and Dr. Detlev W. Bronk, Chairman of the National Research Council, have accepted appointments as members of the Scientific Advisory Committee of Brookhaven National Laboratory, Long Island, New York.

### Louisville Research Institute Dedicated

The University of Louisville Institute of Industrial Research (affiliated with Speed Scientific School) dedicated its new research building on June 12, 1947.

### Herman F. Mark Accepts Belgian Chair

Herman F. Mark, Director of the Institute of Polymer Research and Professor of Organic Chemistry, Polytechnic Institute of Brooklyn, has been invited to accept the Chair Franccki at the University of Liége, Belgium, for the academic year 1947–1948. The guest professorship involves a six-month lecture tour of several Belgian universities.

### *Spectrochimica Acta* Re-Established

The journal *Spectrochimica Acta* has been re-established as an international organ for spectroscopists and spectrochemists, to be published in the Vatican under the editorship of Dr. Alois Gatterer, assisted by a group of coworkers in other countries, as follows: Dr. R. Breckpot of Belgium, Dr. H. Kaiser of Germany, Dr. E. Van Someren of England, and Dr. Lester W. Strock of America. It is expected that the first number of the journal will be ready for press by August 1, and that from six to eight numbers a year will be issued thereafter. Papers in English from America should be submitted to Dr. Strock at 21 Madison Avenue, Saratoga Springs, New York. The official languages of *Spectrochimica Acta* are English, French, and German, and papers will also be accepted in Italian and Spanish.

### Second Instrumentation Exhibit

Over ninety manufacturers of instruments and devices for measurement, inspection, testing, and control have already signed up to exhibit their latest developments at the coming second annual Instrument Conference and Exhibit at the Stevens Hotel, Chicago, September 8–12, 1947, sponsored by the Instrument Society of America.

### Lehigh to Remodel Two Buildings

Lehigh University is remodelling its physics and chemistry buildings at an approximate cost of $105,000. Work is expected to be completed before the opening of the fall semester. The physics building will be equipped with new laboratories for high voltage x-ray equipment, electronics, electric waves, and optical and spectroscopic work.

### Awards

Otto Stuhlman, Jr., Professor of Physics, University of North Carolina, received the Poteat Award at the 64th annual meeting of the North Carolina Academy of Science, May 9–10, in recognition of his biophysical paper "A Dynamical Analysis of the Movements of the Lobes of the Venus' Flytrap."

Presidential Medals for Merit were presented on May 5 to the following scientists for outstanding contributions to the Navy's war effort: John T. Tate, Chairman of Research and Professor of Physics, University of Minnesota; Frederick V. Hunt, Chairman of the Department of Engineering Sciences, Harvard University; Gaylord P. Harnwell, Chairman of the Department of Physics, University of Pennsylvania; and William V. Houston, President of Rice Institute.

### Detection of Isotopes by Tracer Micrography

A method for the more effective tracing of radioactive isotopes in materials in which they have been intentionally introduced has been developed by L. Marton of the National Bureau of Standards with the cooperation of P. H. Abelson of the Department of Terrestrial Magnetism, Carnegie Institution of Washington. In this procedure, by means of a magnetic focusing arrangement, the radiation given off by a radio-isotope within a sample material is made to form an image of the emitting surface upon a photographic plate. The image may then be used in studying the distribution and concentration of the radioactive element present in the sample.

In many chemical, biological, biochemical, and other fields of research, there is growing application of the method of tracers, in which the isotope of a given element is used as an indicator to tag or label certain groups of atoms so that they may be distinguished from other atoms of the same kind. Identification of tracer elements is at present greatly facilitated through the use of radioactive isotopes, which, because of recent developments in atomic energy, are now available in large quantities and are relatively easily detected through their radiations.

In the well-known method of radio autography a radioisotope is introduced in a biological or other system, and the distribution of that particular element within the system is determined by bringing the sample in close contact with a photographic emulsion. This method lacks resolving power, because, even in case of perfect contact of the sample with the emulsion, the circle of confusion from every point of emission is so great that details less than a tenth of a millimeter are very difficult or impossible to distinguish.

In order to improve the resolution of this tracer method, it was decided to use electron optical image formation for determination of the distribution of a radioactive element within a given sample. This process, which may be called "tracer micrography," is based on the emission of high speed electrons (beta-rays) by many tracer elements and the use of magnetic lens elements for forming an image on a suitable recording surface.

In the absence of any means for correction of the chromatic aberration of electron optical lenses, the first micrographs were limited to those elements that emit electrons of uniform speed. After some attempts with columbium 93, yttrium 87, strontium 85, strontium 87, protactinium 233, and gallium 67, the latter was selected for the initial tests. Gallium chloride, prepared by chemical separation from zinc, was bombarded by heavy hydrogen nuclei in the cyclotron at the Carnegie Institution, and the solution was evaporated drop after drop on a ¼-inch tantalum disk. Radiation emitted from the surface of the disk, upon passing through a magnetic lens consisting of a small iron-clad coil with Armco iron pole pieces, was brought to a focus on a photographic film at a distance of about 3½ inches. An image of the tantalum disk was thus obtained showing radioactive areas. The conditions were selected so that a linear magnification of 2 was produced.

For calibration of the instrument, the photographic film was replaced by a Geiger counter, and the lens current necessary to produce a maximum number of counts in unit time was determined for radiations of varying velocities. This established the focusing current for a given type of radiation.

In preliminary experiments with samples of different concentration and thickness of the radioactive layer, exposure times ranged from 2 to 12 hours according to the age and concentration of the sample and the numerical aperture of the lens. It was found that micrographs with good definition were obtained consistently when the layer was sufficiently thin to avoid considerable self-absorption. The best resolving power attained so far has been about 30 microns.

The simplicity of this method, both in apparatus and technique, is one of its more important features. Vacuum requirements are very moderate, since the mean free path of the electrons is large in comparison with the apparatus dimensions, even at forepump pressure.

Scientists at the National Bureau of Standards expect to obtain further improvements in tracer micrography through after-acceleration of the beta-particles by means of a homogeneous electrostatic field. Such after-acceleration may well result in reduced exposure time and in better resolution due to a reduction in spherical aberration. A further reason for after-acceleration is that chromatic aberration, which is always present, even in sources emitting particles of uniform speed, can be markedly decreased if the accelerating potential is at least comparable in magnitude to the energy of the primary emission.

### Utah Symposium on Sound

In connection with the Utah Centennial Celebration a Symposium on Sound was held at the Salt Lake Tabernacle and the University of Utah on July 21 and 22. Invited papers were presented by eleven academic and industrial experts in the field of sound.

# New Books

## Scientific Progress in the Field of Rubber and Synthetic Elastomers

INITIATED BY THE LATE ELMER O. KRAEMER, EDITED BY H. MARK AND G. S. WHITBY. Pp. 453, 15½×23½ cm. Interscience Publishers, Inc., New York, 1946. Price $7.00.

This book which is volume II of *Advances in Colloid Science* is a sequel to volume I and describes recent advances in the physics of rubber and synthetic elastomers. The work and plans of the late Elmer O. Kraemer have been admirably carried out by the co-authors, H. Mark and G. S. Whitby.

This book consists of ten chapters on various phases of high polymer physics, each written by specialists; and they have all done an excellent job. In addition, a biography of the late Dr. Kraemer is included with a list of his many publications. Complete author and subject indexes are included. The book contains around 450 pages and over 800 references.

The first chapter or introduction gives some well-known facts concerning the origin, chemical nature, and some of the properties of the cured polymers of both natural rubber and the more common synthetic elastomers.

The second chapter, a real treatise on the subject of second-order transitions, is extremely well written and contains over 125 references. Tables on second-order transition temperatures and brittle points are given for over thirty different plastics, rubbers, and synthetic rubbers. The theory of second-order transitions is elaborated upon with particular reference to the theory of segment rotation. Experimental results, together with factors influencing second-order transitions, make this section of particular value to workers in the high polymer field.

Chapter 3 is an excellent résumé of the past and most recent information and theories on crystallization phenomena in synthetic and natural rubbers. This section contains a complete bibliography of 150 references.

The fourth chapter goes into a study of crystal and molecular structure of high polymers by means of x-ray diffraction methods. There is some duplication of information in this and the previous chapter; however, it is not objectional, as crystallization phenomena are followed in the first instance primarily by dilatometer and other methods, and in the second instance primarily by x-ray diffraction methods.

There are also valuable chapters on thermodynamics of rubber solutions and gels, viscosity of high polymers, kinetic theory of rubber elasticity, vulcanization, rubber photogels, and photo-vulcanizates. The volume fittingly concludes with a chapter on reinforcing agents in rubber.

Some of the newer organic plastics and resins as vulcanizing agents and as reinforcing agents for rubber and synthetic rubbers are not covered in this volume.

The authors are to be congratulated on the general make-up and quality of the chapters on the timely subjects which are covered in this book.

Duplication of information is surprisingly low in a volume covering such a diversity of inter-related subjects by such a large number of authors.

This book will be a welcome library addition to all those interested in the field of high polymer physics.

R. S. Havenhill
*St. Joseph Lead Company, Monaca, Pennsylvania*

## Electron and Nuclear Counters—Theory and Use

By Serge A. Korff. Pp. 212, 22×14½ cm. D. Van Nostrand Company, Inc., New York, 1946.

In his book, *Electron and Nuclear Counters - Theory and Use*, Korff has gone a long way towards removing the aura of "magic" from detectors used in the field of nuclear physics. The student desiring to enter this field has had in the past practically no recourse but to learn peculiarities of nuclear detectors the hard way, namely, the trial and error method of making and using them. This same student now has available in this very readable book a fairly complete qualitative account of the operation and construction of ionization chambers, proportional counters, and Geiger counters.

The first chapter is devoted to the history and general operation of particle detectors. Korff points out that the main differences between the ionization chamber, proportional counter, and Geiger counter is the voltage region in which they are operated. The next three chapters give a more detailed account of the operation of these detectors. The last of these chapters, namely, the one covering the Geiger counter, is by far the most detailed and complete since this is the author's special field. Even though they have much theory and operation in common, Korff discusses separately the non-self-quenching counter and the self-quenching counter. Expecially interesting is the discussion of the quenching process caused by the presence of certain polyatomic gases in a Geiger counter. The presence of spurious counts or a continuous discharge in a Geiger counter is interpreted as being caused by the emission of photoelectrons and/or secondary electrons from the outer electrode because of photons or positive ions formed in the discharge striking this electrode. If certain polyatomic gases are present they are found to absorb the photons and neutralize the positive ions of the counter gas, the excitation energy of the polyatomic molecule then being dissipated harmlessly through the process of predissociation.

The last three chapters cover, respectively, the constructional aspects of nuclear counters, statistics of counting random events, and finally the auxiliary electronic circuits involved, such as quenching circuits, scaling circuits, coincidence circuits, and linear amplifiers.

The book contains several erroneous and conflicting statements and at least one incorrect circuit diagram. At one point the author speaks of reducing the alpha-particle background of an ionization chamber through the use of a positively charged grid which "would repel any alpha-particles approaching it and drive them back into the wall." The expression for the percentage of missed counts has been developed in Chapter 6 on the basis of very naive assumptions and is valid only for recovery times small compared to the average time between counts. The Neher-Pickering circuit as presented is obviously not capable of quenching. In several cases the circuits given in the book are not the most up-to-date ones used for that particular purpose.

This book, the reviewer feels, has provided the newcomer with a good way of getting acquainted with the field but does not fill the need for a critical and detailed review of the present status of our knowledge of counter mechanisms. A worker in this field will still have to refer to the literature for a complete and quantitative account.

Erwin F. Shrader
*Case Institute of Technology*

## Colloids, Their Properties and Applications

By A. G. Ward. Pp. 133, Interscience Publishers, Inc., New York, 1946; Blackie and Son, Ltd., Glasgow.

A book on colloid chemistry of this limited size can obviously serve, as the author intended it should, only as an elementary account of recent progress in colloid science, and as an introduction of the subject to those who are unfamiliar with it. In general it accomplishes these aims in a satisfactory manner, and can be recommended highly.

The book has three major subdivisions: I. The Nature of the Colloidal State; II. The Colloidal Systems; III. Colloids in Industry and Living Matter. The treatment throughout is descriptive in nature. The author has tried to make the field of colloidal dimensions intelligible to the layman by giving all dimensions first in microns and then, in parenthesis, on a scale one hundred million fold enlarged. Thus a particle of tobacco mosaic virus is said to have a length of 0.5 μ (50 meters). This is rather disconcerting to one who opens the book at random, unaware of the convention adopted.

The author points out that the usual tendency to explain trends in scientific progress on the basis of social and economic needs of the community is not valid in the colloid field. The practice of colloid chemistry is one of man's oldest endeavors in such realms as textiles, pottery soils, paints, etc. Colloid science has lagged behind practice mainly because the tools which were needed to explore the colloidal domain are all fairly new. Hence he places special emphasis on these new tools.

It was a pleasure to read this book and see a broad field free from the complex technical and mathematical details which one needs for active work in colloid chemistry.

R. F. BOYER
*The Dow Chemical Company*

# New Booklets

The National Research Council of the National Academy of Sciences, Prevention of Deterioration Center, Room 204, 2101 Constitution Avenue, Washington, D. C., can now offer the *Prevention of Deterioration Abstracts* on a yearly subscription basis. These abstracts are set up under the following headings: electrical and electronic equipment; finished assemblies; fungicides; lacquers, paints, and varnishes; leather; lubricants; metals; microorganisms; optical instruments; packaging; paper; plastics; resins, rubbers, and waxes; storage, textiles; and wood. Items abstracted include journal articles, patents, specifications, unpublished reports prepared by various Army, Navy, and other governmental groups, and unpublished British, Australian, and Canadian reports.

There will be approximately 1500 pages of the abstracts per year. The individual abstracts are in loose leaf form, so that they may be arranged in manner desired by the individual receiving them. Throughout the calendar year, all the abstracts classified under any one heading will be numbered consecutively.

Comments made by the personnel of the Prevention of Deterioration Center are added to many of these abstracts. In these comments attempts are made to relate a specific report with other relevant ones, to evaluate reports, or to make suggestions concerning further needed research.

The price, which includes two binders and index guides, will be $37.50 per year. Two binders are required for one year's subscription. The fiscal year will be from July 1 to June 30. For the year 1946 back issues will be supplied, since these abstracts started in April 1946.

Dow Corning Corporation, Midland, Michigan, has published a new silicone data sheet entitled *Specifications for Rewinding Motors with Silicone Insulation.* 2 pages, free on request.

The Gaertner Scientific Corporation, 1201 Wrightwood Avenue, Chicago 14, Illinois, has announced publication of two bulletins: 157–74, *Divided Circle Spectrometers and Accessories,* 20 pages; 158–74, *Wavelength Spectrometers and Monochromators,* 12 pages. Available free on request.

The Tube Department of Radio Corporation of America, Harrison, New Jersey, has brought out a revised edition of the *Quick-Reference Chart on Miniature Tubes.* It is form MNT-30A. 4 pages.

North American Philips Company, Inc., 100 East 42d Street, New York 17, New York, has published a two-page chart (R-1066) entitled *Basic Characteristics of Useful Industrial Laboratory Instruments.* Data are presented in a convenient form to facilitate hanging on the wall for ready reference. Free on request.

Littelfuse, Inc., 4757 Ravenswood Avenue, Chicago 40, Illinois, has recently published its first completely new catalog since the beginning of the war. It is a manual of progress in the field of circuit protection, including a brief historical survey of fusing and the announcement of many new developments in the company's line of fuses, fuse mountings, and circuit indicators. The catalog's number is 9.

Dewey and Almy Chemical Company, Boston 40, Massachusetts, has issued a new catalog of *Darex* meteorological balloons, containing many new performance and data charts calibrated in both the English and metric systems. 30 pages. Free on request.

RCA Application Note for May 15, 1947, published by the Tube Department of Radio Corporation of America, Harrison, New Jersey, is entitled *Use of the 2E24 and 2E26 at 162 Megacycles.* 8 pages, available on request.

Interscience Publishers, Inc., 215 Fourth Avenue, New York 3, New York, has published a 50-page catalog of its recent releases in the fields of chemistry, physics, mathematics, engineering, and medicine.

Eastman Kodak Company, Rochester 4, New York, has issued a revised data book on *Infrared and Ultraviolet Photography.* Priced at 25 cents a copy, it is available through all Kodak dealers.

Journal
of
Applied Physics

Volume 18, Number 9 September, 1947

# Microwave Dielectric Measurements

T. W. Dakin and C. N. Works
*Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania*

(Received March 7, 1947)

The method of measurement of the dielectric properties of solid dielectrics described by S. Roberts and A. von Hippel has been applied with 3-cm wave-length waves in a rectangular wave guide. In this method the dielectric sample is placed at a short circuited end of the transmission line, and the dielectric properties of the sample, calculated from the position of a minimum of the standing wave and the ratio of the minimum field strength to the maximum field strength of the standing wave. A simplified procedure for calculating the dielectric properties from the measurements is presented. Values are given of the dielectric properties of some common plastic and ceramic materials at 10-cm wave-length and 3-cm wave-length, which were measured with this method.

## INTRODUCTION

At microwave frequencies the dielectric properties of a material can be obtained accurately only by measuring the wave-length in the material or its reflecting power and the attenuation of the wave in the material. There are many experimental ways in which the problem can be attacked. A summary of the experimental work in this field up to 1927 is given by Romanoff.[1] His summary includes the pioneering work of Drude,[2] who proposed two methods, the first of which consisted of measuring the wave-length directly in the dielectric media surrounding a resonant section of a pair of lecher wires, and the second of which consisted of measuring the capacitance of a small capacitor containing the dielectric and closely attached to the resonant section of lecher wires. Drude's second method contains inherent errors, when applied at very short wave-lengths, because of the inaccuracy of considering a capacitor as a lumped capacitance at these wave-lengths. Drude's second method has, however, been used by many investigators[3] even up to recent years. Modifications and improvements of Drude's first method, including experiments where only a part of the resonant section of a transmission line is filled with dielectric, have been made by many investigators[4] since the summary of Romanoff. In more recent work, enclosed coaxial or hollow

---

[1] W. Romanoff, *Handbuch der Physik* (Julius Springer Verlag, Berlin, 1927), Vol. XV, p. 491.

[2] Paul Drude, Zeits. f. physik. Chemie **23**, 267 (1897). Ann. Physik und Chemie **55**, 633 (1895); *ibid.* **61**, 466 (1897).

[3] W. Hemple, Elekt. Nachr.-Techn. **14**, 33 (1937); J. W. Miller and B. Salzberg, R.C.A. Rev. **3**, 486 (1938); R. Odenwald, Ann. d. Physik **35**, 690 (1939); H. Slatis, Ann. d. Physik **36**, 397 (1939); C. Klages, Physik. Zeits. **43**, 151 (1942).

[4] M. Seeberger, Ann. d. Physik **16**, 77 (1933); H. Kaufmann, Hoch: tech. u. Elek: akus. **53**, 61 (1939); J.; Malsch, Ann. d. Physik **12**, 865 (1932); *ibid.* **19**, 707 (1934); *ibid.* **20**, 33 (1934); M. Ardenne, Otto Groos, and G. Otterbein, Physik. Zeits. **37**, 533 (1936); R. King, Rev. Sci. Inst. **8**, 201 (1937); K. Slevogt, Ann. d. Physik **36**, 141 (1939).

transmission lines have been used.[5] Optical methods where the transmission through slabs of the dielectric and reflection from the dielectric surface are measured have also been used[6] with microwaves. Direct measurement by probes of standing waves set up within liquid dielectrics and in front of dielectric surfaces by reflection have also been used to obtain dielectric properties.[7] Resonant-cavity methods where all or part of the cavity is filled with dielectric are an extension of the resonant-transmission line idea and these have been used at microwave frequencies.[8]

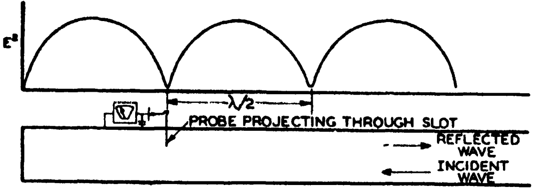


FIGURE 1a (NO DIELECTRIC)

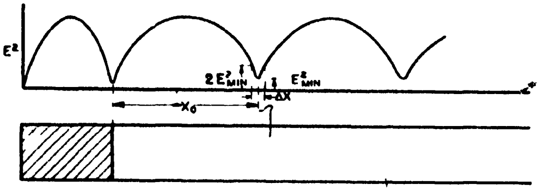


FIGURE 1b (HALF WAVE-LENGTH SECTION OF DIELECTRIC)

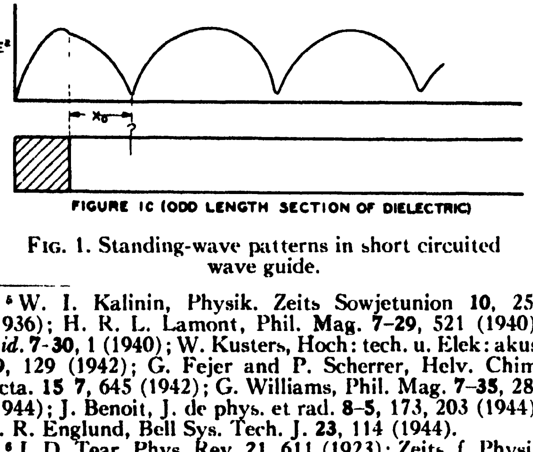


FIGURE 1c (ODD LENGTH SECTION OF DIELECTRIC)

Fig. 1. Standing-wave patterns in short circuited wave guide.

## MEASUREMENT TECHNIQUE AND CALCULATIONS

The method using standing-wave measurements in the transmission line in front of the dielectric sample proposed by S. Roberts and A. von Hippel[9] has been developed for the measurement of solid dielectrics to a more practical state than any other method in recent years. This is the method which is used and discussed in this paper. This method consists of reflecting the wave at normal incidence from a slab of the dielectric which is placed against a perfectly reflecting surface. The process of reflection sets up standing waves in the space region in front of the sample as a result of the superposition of the incident and reflected waves (See Figs. 1a, 1b, 1c). The electric field direction is maintained at right angles to the direction of propagation of the wave (TE wave).

In actual practice the wave and the dielectric sample are restricted to an enclosed hollow or coaxial wave guide, although this is not in principle a necessary restriction. The same equations are valid in principle for a measurement using lecher wires or free space with a parallel beam of radiation, although it is more difficult in practice to do measurements under those conditions.

The separation of the first minimum from the face of the sample will depend on the wave-length of the wave in the sample and its thickness, since the first minimum will be an integer number of half-waves from the reflecting surface behind the sample (See Fig. 1b). In Figs. 1b and 1c the shortening of the wave-length in the sample is shown. Insertion of the dielectric shifts the minima of the standing wave toward the end. The separation of the first minimum of the standing wave from the surface of the sample is a measure of the dielectric constant.

If there is any attenuation in the dielectric, there will be a reduction in the standing-wave ratio $E_{\max}/E_{\min}$, since all the incident wave intensity would not be reflected. There is, of course, a finite standing-wave ratio with no dielectric present, because of the slight attenuation by the copper- or silver-walled transmission line. The measured value of $E_{\max}/E_{\min}$ is mathematically

---

[5] W. I. Kalinin, Physik. Zeits Sowjetunion **10**, 257 (1936); H. R. L. Lamont, Phil. Mag. **7-29**, 521 (1940); *ibid.* **7-30**, 1 (1940); W. Kusters, Hoch: tech. u. Elek: akus. **59**, 129 (1942); G. Fejer and P. Scherrer, Helv. Chim. Acta. **15 7**, 645 (1942); G. Williams, Phil. Mag. **7-35**, 283 (1944); J. Benoit, J. de phys. et rad. **8-5**, 173, 203 (1944); C. R. Englund, Bell Sys. Tech. J. **23**, 114 (1944).

[6] J. D. Tear, Phys. Rev. **21**, 611 (1923); Zeits. f. Physik **113**, 415 (1939); W. Kebbel, Hoch: tech. u. Elek: akus. **53**, 81 (1939), G. Baz, Physik. Zeits. **40**, 394 (1939); L. Palmer and G. Forrester, Proc. Phys. Soc. London, **53**, 479 (1941); L. H. Ford and R. Oliver, Proc. Phys. Soc. London **58**, 265 (1946).

[7] M. Velasco and G. L. Hutchinson, Proc. Phys. Soc. London **51**, 689 (1939); A. N. Soos, Comptes rendus Acad. Sci. U.R.S.S. **33**, 310 (1941); W. P. Conner and C. P. Smyth, J. Am. Chem. Soc. **65**, 382 (1943).

[8] F. Borguis, Naturwiss. **29**, 516 (1941); C. N. Works, T. W. Dakin, and F. W. Boggs, Trans. A. I. E. E. **63**, 1092 (1944); D. L. Hollaway, J. Inst. Eng. Aust. **21**, 79 (1940); W. R. MacLean, J. App. Phys. **17**, 558 (1946).

[9] S. Roberts and A. von Hippel, J. App. Phys. **17**, 610 (1946).

related to the attenuation constant and the dielectric constant of the slab of dielectric by the relation (1) which was derived by S. Roberts and A. von Hippel.[9]

$$\frac{\dfrac{E_{\mathrm{min}}}{E_{\mathrm{max}}}-i\tan\dfrac{360x_0}{\lambda_g}}{1-i\dfrac{E_{\mathrm{min}}}{E_{\mathrm{max}}}\,\dfrac{\tan 360x_0}{\lambda_g}}\cdot\frac{\lambda_g}{2\pi d i}=\frac{\tanh\gamma_2 d}{\gamma_2 d}. \tag{1}$$

Where $x_0$ is the distance from the first electric field strength minimum of the standing wave to the face of the sample, $d$ is the thickness of the sample, and $\lambda_g$ is the wave-length along the wave guide outside the sample. $E_{\mathrm{min}}/E_{\mathrm{max}}$ is the ratio of the electric field strength at an adjacent maximum point. $\gamma_2$ is the propagation constant of the wave in the wave guide and within the sample.

The most general procedure is to solve the above complex transcendental equation for $\gamma_2 d$ after the measured values of $E_{\mathrm{min}}/E_{\mathrm{max}}$, $\lambda_g$, $d$, and $x_0$ are obtained. The solution of this equation can be obtained directly only by the use of charts of the complex function: $\tanh\gamma_2 d/\gamma_2 d$. Such charts have been prepared by Professor von Hippel and co-workers.[10] A convergent series solution has also been obtained by them for the case where $360x_0/\lambda_g$ is near to 90° or an odd multiple of 90°, and $E_{\mathrm{min}}/E_{\mathrm{max}}$ is small. But this direct method of solving the complex transcendental equation is very laborious.

A simpler solution, usually of sufficient accuracy, can be obtained when the value of the dissipation factor of the dielectric is less than 0.1 by separating Eq. (1) into its real and imaginary parts and equating them separately, remembering that $\gamma_2=\alpha_2+\beta_2 i$. Equation of the real parts gives:

$$\frac{-\lambda_g\tan\dfrac{360x_0}{\lambda_g}\left(1-\dfrac{E^2_{\mathrm{min}}}{E^2_{\mathrm{max}}}\right)}{2\pi d\left(1+\dfrac{E^2_{\mathrm{min}}}{E^2_{\mathrm{max}}}\tan^2\dfrac{360x_0}{\lambda_g}\right)}=\frac{\beta_2 d\tan\beta_2 d(1-\tanh^2\alpha_2 d)+\alpha_2 d(1+\tan\beta_2 d)\tanh\alpha_2 d}{(\alpha_2{}^2d^2+\beta_2{}^2d^2)(1+\tanh^2\alpha_2 d\tan^2\beta_2 d)}. \tag{2}$$

When $\alpha_2 d$ and $E_{\mathrm{min}}/E_{\mathrm{max}}$ are small, the above equation reduces to:

$$\frac{-\lambda_g\tan\dfrac{360x_0}{\lambda_g}}{2\pi d}=\frac{\tan\beta_2 d}{\beta_2 d}. \tag{3}$$

This is an equation involving no complex numbers which may be readily solved for $\beta_2 d$ by reference to a table of $(\tan\theta)/\theta$ for radian arguments. Such an abbreviated table exists in Jahnke Emde's ***Funktionentafeln***. A more complete table has been prepared at the Westinghouse Research Laboratory, and another by the Laboratory for Insulation Research at MIT.[11] The values on the left side of Eq. (3) are measured quantities, and after $\beta_2 d$ has been obtained the dielectric constant can be calculated from the following equation:

$$\epsilon'=\frac{1/\lambda_c{}^2+\dfrac{(\beta_2 d)^2}{(2\pi d)^2}}{1/\lambda_c{}^2+1/\lambda_g{}^2}, \tag{4}$$

where $\epsilon'$ is the dielectric constant, $\lambda_c$ is the cut off wave-length of the wave guide. If coaxial wave guide is used, $1/\lambda_c{}^2$ is zero. The above equation is accurate only for materials having dissipation factors less than 0.1. The solution of Eq. (3) for $\beta_2 d$ is multi-valued and will give several values of $\epsilon'$. If the correct one cannot be selected by previous knowledge of the material, a second measurement is necessary using a sample of different length, $d$.

An expression can now be obtained for the attenuation constant by equating the imaginary

---

[11] Report IX by W. B. Westphal, Laboratory for Insulation Research, MIT Contract OEMsr-191 N.D.R.C.

[10] Report III by A. von Hippel, D. G. Jelatis, and W. B. Westphal, Laboratory for Insulation Research, MIT Contract OEMsr 191, N.D.R.C.

parts of Eq. (1).

$$\frac{-\lambda_g\dfrac{E_{\min}}{E_{\max}}\left(1+\tan^2\dfrac{360x_0}{\lambda_g}\right)}{2\pi d\left(1+\dfrac{E^2_{\min}}{E^2_{\max}}\tan^2\dfrac{360x_0}{\lambda_g}\right)} = \frac{\alpha_2 d\tan\beta_2 d-\alpha_2 d\tanh^2\alpha_2 d\tan\beta_2 d-\beta_2 d\tanh\alpha_2 d-\beta_2 d\tanh\alpha_2 d\tan^2\beta_2 d}{(1+\tanh^2\alpha_2 d\tan^2\beta_2 d)(\alpha_2^2d^2+\beta_2^2d^2)}. \quad (5)$$

When $\alpha_2$ and $E_{\min}/E_{\max}$ are small, Eq. (5) reduces to:

$$\alpha_2 d \cong \frac{\beta_2^2d^2\lambda_g}{2\pi d}\cdot\frac{E_{\min}}{E_{\max}}\cdot\frac{\left(1+\tan^2\dfrac{360x_0}{\lambda_g}\right)}{\beta_2 d(1+\tan^2\beta_2 d)-\tan\beta_2 d}. \quad (6)$$

Roberts and von Hippel[9] have shown that $E_{\min}/E_{\max}=\pi\Delta x/\lambda g$ where $\Delta x$ is the width of the minimum of the standing wave to the double power points. After making this substitution, and putting in an equivalent value for $\alpha_2$, Eq. (6) gives an analytical expression for the dissipation factor:

$$\tan\delta_2=\frac{\Delta x}{d}\left[\frac{(1/\lambda_c^2+1/\lambda_g^2)-1/\lambda_c^2\epsilon'}{(1/\lambda_c^2+1/\lambda_g^2)}\right] \times\frac{\beta_2 d\left(1+\tan^2\dfrac{360x_0}{\lambda_g}\right)}{\beta_2 d(1+\tan^2\beta_2 d)-\tan\beta_2 d}. \quad (7)$$

In the above equation the term within the brackets reduces to unity for coaxial wave guide, since $1/\lambda_c^2$ is zero in that case. In carrying out calculations it is first necessary to obtain $\beta_2 d$ and $\tan\beta_2 d$ from Eq. (3) before the dissipation factor is calculated using Eq. (7).

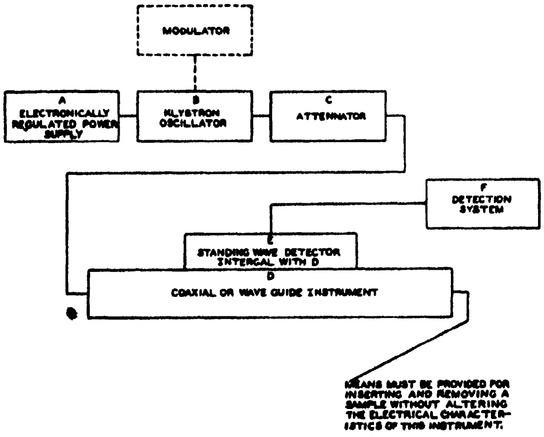

MODULATOR
A
ELECTRONICALLY REGULATED POWER SUPPLY
B
KLYSTRON OSCILLATOR
C
ATTENNATOR
F
DETECTION SYSTEM
E
STANDING WAVE DETECTOR INTERGAL WITH D
D
COAXIAL OR WAVE GUIDE INSTRUMENT
MEANS MUST BE PROVIDED FOR INSERTING AND REMOVING A SAMPLE WITHOUT ALTERING THE ELECTRICAL CHARACTERISTICS OF THIS INSTRUMENT.

Fig. 2.

Since there is always some dissipation in the walls of the wave guide, a measurement of $E_{\min}/E_{\max}$ includes those losses together with the dielectric losses. These wave guide losses must be subtracted. The proper procedure is first to correct the measured value of $E_{\min}/E_{\max}$ for the loss due to the wave guide in front of the sample, between the sample face and the measured minimum. Then the total dissipation factor of the wave guide and sample are calculated from Eq. (7), and the dissipation factor of the wave guide subtracted from the result. The dissipation factor of the wave guide can be obtained from Eq. (7) after measuring the value of $\Delta x$ with the wave guide empty. In this case $\epsilon'$ would be unity, $d$ would be the distance from the measured minimum to the shorted end of the wave guide, and both $\tan(360x_0/\lambda_g)$ and $\tan\beta_2 d$ would be very large. Then Eq. (7) gives for the dissipation factor of the wave guide:

$$\tan\delta_w=\frac{\Delta x}{d}\,\frac{1}{1+\lambda_g^2/\lambda_c^2}. \quad (8)$$

Since the approximations involved in going from Eqs. (2) to (3) and from (5) to (6) are not readily obvious except in the limiting cases of very large $\tan(360x_0/\lambda_g)$ and $\tan\beta_2 d$ or of $\tan(360x_0/\lambda_g)$ and $\tan\beta_2 d$ less than unity, a numerical check of the validity of the approximations was made by substituting in typical measured values of $\Delta x$, $d$ and $(360x_0/\lambda_g)$. It was verified by numerical calculation that Eq. (3) is accurate to about ±1 percent in the dielectric constant for dissipation factors less than 0.1. This numerical check was made for a wide range of values of $(360x_0/\lambda_0)$, varying from 0° to 120° and also a wide range of values of $\Delta x$ and $d$ by comparing values of $\epsilon'$ and $\tan\delta$ obtained from Eqs. (3), (4), and (7) with those obtained from a direct graphical solution of Eq. (1) for $\gamma_2$. The accuracy of the approximation increases with decreasing $\tan\delta$. Similar accuracy is also obtained in the calculation of $\tan\delta$

## EXPERIMENTAL

The complete electrical system is shown in schematic form in the block diagram, Fig. 2. Figures 3 and 4 show the standing-wave detector and sample holding section of the instrument. The cross section of the guide and the sample is 0.900" by 0.400". Typical samples vary in length from about ⅛" for high loss and high dielectric constant materials to about 1½" for low loss and low dielectric constant materials.

The position of the probe must be known with great precision, so a Gaertner comparator, No. 1658, 0–100 mm, which reads accurately to 0.001 mm, was used to carry the probe. The probe must have sufficient length to give sufficient signal at the minimum of the wave pattern; its presence should not disturb the field; it must be shielded so that it does not produce disturbing radiation.

The steel brackets shown as $E$ and $F$ in Fig. 3 enabled a careful adjustment of the comparator so that the probe would move parallel (within ±0.0001") with the bottom of the wave guide. The wave guide was milled from two pieces of brass with the inside silver plated. The dimensions of the guide were held to a tolerance of ±0.0001" with parallel top and bottom. The 13-cm slot which accommodates the probe was made narrow (0.070") and deep (0.175") to reduce the radiation from it. This choice of slot dimensions makes it possible to place one's fingers on the slot without changing the response of the detector.

Details of the probe assembly are shown in Fig. 4. The probe which extends into the guide is a silver-plated steel wire of 0.012" diameter, and its extent of penetration into the guide is adjustable. The probe is tuned with an adjustable choke so that maximum energy will be transferred from the probe to the detector.

The wave guide instrument has been used both tuned and untuned. When operating untuned, a tapered non-reflecting attenuator is placed between the end where the power is introduced and the end where measurements are made. This requires the introduction of more energy for the same detector current than in the case where the guide is tuned with a shorting plunger at the end opposite where the measurement is made.

Two different reflex klystron tubes (WE 721A and Sperry 2K25) operating in the vicinity of 10,000 megacycles have been used with the wave-guide instrument.

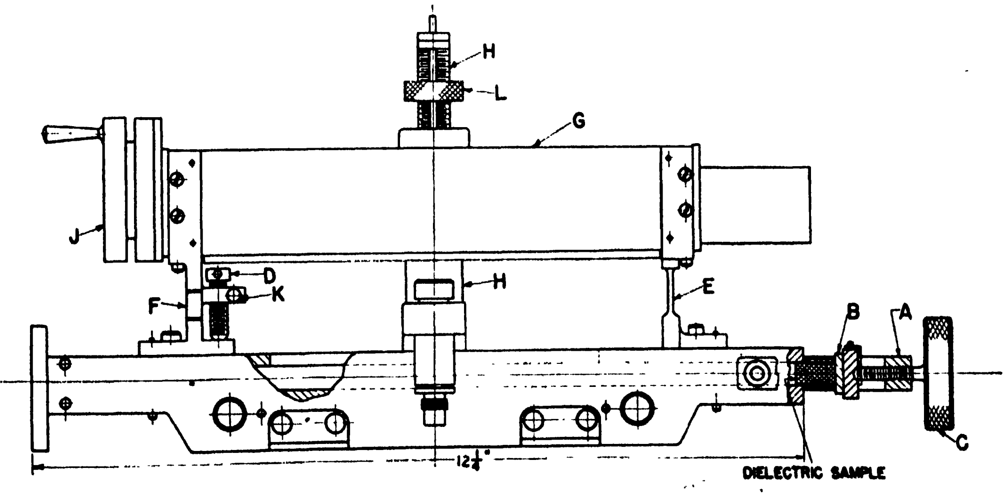


FIG. 3. The wave-guide instrument for dielectric measurements at 3 cm.

The power supplies for these tubes are of the conventional electronic voltage-regulated type.[12] The reflector voltage is pulse modulated by a square wave generator operating at an audiofrequency to amplitude modulate the microwave intensity. But where an automatic frequency control system was used to stabilize further the frequency of the oscillator tube for increased accuracy in the measurement of low loss materials, the microwave intensity was usually not modulated and a d.c. detector was used.

The crystal d.c. current from the 1N21 crystal detector shown in Fig. 4 was assumed to be proportional to the square of the radiofrequency electric field in the wave guide for low currents at which it is used. Galvanometers have been used as crystal current meters when the microwaves are unmodulated. The galvanometer should have a sensitivity of at least 0.002 micro ampere per mm to give a deflection that is readable with sufficient accuracy for measurements at the minimum voltage region.

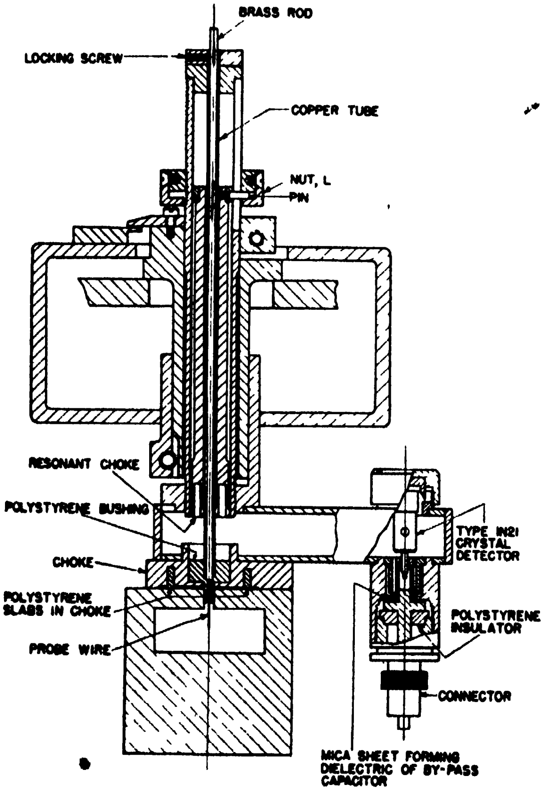


FIG. 4. Cross section through probe assembly and wave guide.

The rapid response of an a.c. amplifier makes it preferred over a galvanometer as a crystal-current indicator. When the microwave power is pulse modulated at an audiofrequency, a sensitive tuned audiofrequency amplifier is used with a panel output meter. The amplifier must be linear so that the output meter reading is always proportional to crystal current. Use of a reasonably high audiofrequency-tuned amplifier also has the advantage of largely eliminating spurious noise developed in the crystal.

## SOURCES OF ERRORS IN THE MEASUREMENTS

### (1) Errors Due to Change in the Frequency of the Oscillator

Changes in frequency may result during the measurement as a result of too close coupling between the oscillator tube and the measuring wave guide, or because of inadequate regulation of the electrode voltages on the tube. This, of course, results in changes in position of the voltage minima in the wave guide.

### (2) Errors Due to Excessive Projection of the Probe into the Wave Guide

If the probe projects too far into the wave guide, it causes an error in both the position of the minimum and the value of $\Delta x$. If the detecting system and the standing wave detector are adjusted for maximum sensitivity, the required projection of the probe to obtain a satisfactory deflection of the indicating instrument should be such that no error is introduced. When very high loss samples are in the wave guide, it is well to check against this source of error by increasing or decreasing the projection of the probe into the wave guide by about 1/64 inch and again measuring the location of the minimum and the value of $\Delta x$. No change should be observed.

### (3) Errors Due to Frequency Modulation

It has been found that appreciable frequency modulation of the source causes the value of $\Delta x$ to be increased. This effect is analyzed by M. G. Haugen and W. B. Westphal.[13]

---

[12] F. E. Terman, *Radio Engineers Handbook* (McGraw-Hill Book Company, Inc., New York, 1943), p. 614.

[13] M. G. Haugen and W. B. Loss, "The design of equipment for measurement of dielectric constant and loss with

TABLE I.

| Type of wave guide | Max. dimension for lowest TE mode only** | Min. dimension for lowest TE mode** | $\lambda_c$ | Wave-length in sample $\lambda_s$ | Characteristic impedance |
|---|---|---|---|---|---|
| Coax | | | | | |
| $r_0$ = outer radius | $r_0 - r_1 \simeq \frac{\lambda}{2}$ | none | $\infty$ | $^*\lambda_a/(\epsilon')^{\frac{1}{2}}$ | $\frac{^*138}{(\epsilon')^{\frac{1}{2}}} \log_{10}\frac{r_0}{r_1}\left(\frac{1}{1-i\tan\delta}\right)^{\frac{1}{2}}$ |
| $r_1$ = inner radius | $r_0 + r_1 \simeq \frac{\lambda}{\pi}$ | | | | |
| Rectangular | | | | | |
| $a$ = smaller width | $\frac{1}{a^2}+\frac{1}{b^2} > \frac{4}{\lambda^2}$ | $b \geq \frac{\lambda}{2}$ | $\lambda_c = 2b$ | $^*\frac{\lambda_c\lambda_a}{(\lambda_c^2\epsilon' - \lambda_a^2)^{\frac{1}{2}}}$ | $^{***}376.6\left(\frac{\mu}{\epsilon'}\right)^{\frac{1}{2}}\frac{\lambda_s}{\lambda_a}\left(1+\frac{i\epsilon''}{2\left(\epsilon'-\frac{\lambda_a^2}{\lambda_c^2}\right)}\right)$ |
| $b$ = greater width | $b < \lambda$ | | | | |
| Cylindrical | | | | | |
| $r$ = radius | $r < \frac{\lambda}{2.61}$ | $r > \frac{\lambda}{3.41}$ | $\lambda_c = 3.41r$ | $^*\frac{\lambda_c\lambda_a}{(\lambda_c^2\epsilon' - \lambda_a^2)^{\frac{1}{2}}}$ | $^{***}376.6\left(\frac{\mu}{\epsilon'}\right)^{\frac{1}{2}}\frac{\lambda_s}{\lambda_a}\left(1+\frac{i\epsilon''}{2\left(\epsilon'-\frac{\lambda_a^2}{\lambda_c^2}\right)}\right)$ |

* $\tan\delta < 0.1$ and neglecting conductor losses.
** Divide $\lambda$ by $(\epsilon')^{\frac{1}{2}}$ where guide is filled with dielectric.
*** Not comparable to characteristic impedance of coax ($\mu = 1$ for most dielectrics).
$\lambda_a$ = free space wave-length.

However, if the wave form of the square wave signal used to modulate the oscillator is properly adjusted so that it reaches maximum value in a very short time and has a flat top or, in other words, approaches an ideal square wave closely, the frequency modulation introduced in this way will be so small that the error in the value of $\Delta x$ will be less than the precision with which this quantity may be measured, because of mechanical limitations in the design of the standing wave detector.

### (4) Errors Due to Faulty Tuning

Changes in the apparent value of the width of the minima have been noted when the wave guide is tuned slightly off resonance. This does not occur when operating with a non-reflecting attenuator and an untuned section of the guide.

### (5) Higher Transmission Modes in the Dielectric

It is possible to have higher modes than the lowest TE mode within the dielectric at the end of the wave guide, even though the wave guide is beyond cut off for those modes in that part of the guide filled with air. The existence of other modes within the dielectric will cause a spurious reflection at the surface of the dielectric and lead to erroneous results in the dielectric constant and dissipation factor.

From the relations in Column 2 of Table I, it can be seen that when a dielectric is present, the wave-length should be longer if only the lowest TE mode is desired. Probably resonant lengths of dielectric samples, i.e., ones which are a multiple of $\frac{1}{4}$ or $\frac{1}{2}$ wave-length (in the sample) for the higher mode, should be avoided, for these lengths are more likely to give the higher modes. The values of $\lambda$ given by the expressions in Column 2 of Table I give the cut off wave-lengths for the next higher modes. From these the wave-length in the guide in the dielectric sample can be calculated for the higher mode of transmission. The effect of higher modes on the accuracy of measurements has not been thoroughly investigated. Care should be taken especially with high dielectric constant materials where the possibility of higher modes is increased. Even with the coax instrument, the possibility of higher modes is prevalent.

## MEASUREMENTS—DATA

Dielectric measurements at 3-cm wave-length and 10-cm wave-length have been made on a variety of plastic and ceramic materials during a

---

standing waves in wave guides," Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, National Defense Research Committee, Contract OEMsr-191, Report XII.

TABLE II. Dielectric properties at 3-cm wave-length.

| Material | $\epsilon'$ | tanδ | $\alpha$, Nepers per meter, coax or free space* | D.B. per meter, coax or free space* | $N$, refractive index | Power-reflection coefficient normal incidence |
|---|---|---|---|---|---|---|
| Polyethylene | 2.29 | .00032 | 0.051 | .44 | 1.514 | .042 |
| Polystyrene | 2.43 | .00036 | 0.060 | .52 | 1.59 | .052 |
| Teflon | 2.05 | .00025 | 0.038 | .33 | 1.431 | .031 |
| NaCl crystal | 5.78 | .0002 | 0.05 | .43 | 2.40 | .17 |
| Insanol** | 7.96 | .0042 | 1.24 | 10.8 | 2.82 | .228 |
| Corning glass 704 | 4.79 | 0.0063 | 1.45 | 12.6 | 2.19 | .139 |
| Corning glass 707 | 4.05 | 0.0022 | 0.46 | 4.0 | 2.01 | .113 |
| Micarta 199† | 4.03 | 0.0624 | 13.14 | 114.3 | 2.01 | .113 |
| Micarta 496† | 3.58 | 0.054 | 10.7 | 93. | 1.89 | .095 |
| Micarta 254† | 3.36 | 0.039 | 7.47 | 65. | 1.83 | .086 |
| Micarta 299† | 4.61 | 0.0211 | 4.75 | 41.3 | 2.15 | .133 |
| Micarta 259† | 5.36 | 0.041 | 9.95 | 86.6 | 2.32 | .158 |
| Micarta 273† | 3.56 | 0.0686 | 13.57 | 118.0 | 1.89 | .095 |
| Corex resin | 2.69 | 0.0031 | .53 | 4 6 | 1.64 | .059 |
| Polydichlorstyrene | 2.63 | 0.0005 | .085 | .74 | 1.62 | .056 |
| Nylon | 3.09 | 0.0106 | 1.95 | 17.0 | 1.76 | .076 |
| 993 Silicone | 2.90 | 0.0050 | .89 | 7.7 | 1.70 | .067 |
| Fosterite 80–20 | 2.57 | 0.0048 | .81 | 7.0 | 1.60 | .053 |
| Fosterite 35–65 | 2.58 | 0.0144 | 2.43 | 21.1 | 1.61 | .055 |
| Molded Micarta 192a | 4.34 | 0.0093 | 2.03 | 17.6 | 2.08 | .123 |
| Kraft board*** | 3.55 | .121 | 23.8 | 207. | 1.88 | .093 |
| Kraft board, dry | 2.99 | .05 | 9.1 | 79. | 1.73 | .072 |
| Dense Kraft board*** | 3.82 | .122 | 24.9 | 217. | 1.95 | .104 |
| Dense Kraft board, dry | 3.20 | .05 | 9.4 | 82. | 1.79 | .080 |
| Ivory, walrus tusk | 6.98 | .0934 | 25.8 | 225. | 2.64 | .204 |

* Divide values in this column by $(1-\lambda^2/\epsilon'\lambda_c^2)^{1/2}$ for attenuation in wave guide.

** Westinghouse product similar to Mycalex.

*** At 30 percent humidity.

† Micarta 199 is a chopped cotton cloth filled phenol formaldehyde plastic. Micarta 273 is a fine weave cotton cloth phenol formaldehyde resin laminate. Micarta 259 is a glass cloth melamine formaldehyde resin laminate. Micarta 254 is a Kraft paper cresol formaldehyde resin laminate. Micarta 299 is a glass cloth cresol formaldehyde resin laminate. Micarta 496 is a medium weave cotton cloth phenol formaldehyde resin laminate.

TABLE IIa. Dielectric properties at 10-cm wave-length.

| Material | $\epsilon'$ | tanδ | $\alpha$, Nepers per meter in coax or free space | DB. per meter in coax or free space | $N$, index of refraction | Power reflection coefficient normal incidence |
|---|---|---|---|---|---|---|
| Polyethylene | 2.29 | 0.0002<br>to 0.0006 | 0.0095<br>to 0.028 | 0.083<br>to 0.25 | 1.506 | 0.041 |
| Polystyrene | 2.50 | 0.0003 | 0.015 | 0.13 | 1.58 | 0.050 |
| Insanol | 7.95 | 0.006 | 0.53 | 4.6 | 2.82 | 0.23 |
| Corning glass 704 | 4.80 | 0.004 | 0.275 | 2.4 | 2.19 | 0.14 |
| Corning glass 707 | 3.96 | 0.0013 | 0.081 | 0.70 | 1.99 | 0.11 |
| Corning glass 774 | 4.45 | 0.0051 | 0.34 | 3.0 | 2.11 | 0.13 |
| Micarta 199 | 4.31 | 0.0860 | 5.6 | 49. | 2.08 | 0.12 |
| Micarta 254 | 3.66 | 0.0632 | 3.8 | 33. | 1.91 | 0.10 |
| Micarta 496 | 4.20 | 0.0988 | 63. | 55. | 2.05 | 0.12 |
| Micarta glass filled | 4.29 | 0.0243 | 1.58 | 13.7 | 2.07 | 0.12 |
| Pork meat | 62.6 | 0.30 | 74. | 640. | 7.9 | 0.60 |

period of several years. These results have been reported to interested persons when they requested the information, but have not been contained in any previous report in a tabulation. These data are now collected together with this paper.

The measurements were all made according to the method outlined in this paper, and the dielectric properties calculated according to the scheme outlined in this paper. The results of these measurements are tabulated in Table II along with useful attenuation and reflection constants of these materials for those frequencies.

The values for 10-cm wave-length were obtained on the MIT designed coaxial wave guide instrument (Reference 10) with the same technique in measurement as with the hollow rectangular wave guide.

In some cases, the materials are quite variable in composition, and deviations from the values reported in Table II might be found. In some cases, the dissipation factor and dielectric constant are quite sensitive to moisture. In these cases, the conditions of humidity or dryness of the materials when the samples were measured are given.

## SUMMARY

A method of measurement of the dielectric properties of solids at microwave frequencies has been discussed. A rectangular wave-guide apparatus for performing measurements in the vicinity of 10,000 megacycles has been described in detail. A simplified method of calculating the results of the wave-guide measurements is presented. Included in the paper are values of the dielectric properties of some common plastics and ceramics measured at 10-cm wave-length and 3-cm wave-length.

The authors wish to acknowledge guidance they obtained through the prior and simultaneous work in this field carried out at the laboratory for insulation research at the Massachusetts Institute of Technology under the direction of Professor A. von Hippel. Appreciation is also expressed to members of the Electronics Department of the Westinghouse Research Laboratory, who introduced us to some of the microwave apparatus which we have applied.

# Electron Beam Deflection

## Part II. Applications of the Small-Angle Deflection Theory

R. G. E. Hutter*

(Received February 26, 1947)

The theory developed in Part I is applied to a number of typical deflection fields such as the fields produced by parallel plates, parallel cylinders, semi-infinite co-planar sheets, and bent-plates. Graphs for the magnitude of deflection are given. An electrolytic tank potential-plotting device is described which was used for the determination of the potential distribution of some of the deflection fields.

The theory shows that the deflection is a function of the initial conditions of the electron beam before entering the field and that it is a slightly non-linear function of the field strength. It follows that the fluorescent spot and the pattern produced on the screen of cathode-ray tubes will show distortions. These distortions were calculated for several deflection fields and electron beam shapes; the results are shown graphically.

The problem of the reduction of these distortion effects is discussed.

### A. THE DEFLECTION BY TYPICAL ELECTRIC DEFLECTION FIELDS

ELECTROSTATIC deflection systems are ordinarily designed by experienced personnel. Since no design formulae or charts of reasonable accuracy exist to aid the inexperienced person, he must therefore use the deflection formula for the parallel-plate deflection field without edge effects[1] as a guide. He may then correct the resultant error by adjustment of the deflection potential. This means that deflection amplifiers having a wide gain adjustment must be provided. It is the purpose of this paper to present design charts for a number of typical electric deflection systems.

The paper "Electron Beam Deflection," Part I,[2] contained a mathematical expression for the deflection in terms of the accelerating potential $\phi$ and the axial field-strength distribution $E(z)$ (Part I, Eq. (31)). It is (assuming $x_0=y_0=x_0'=y_0'=0$):

$$d=Y(z_i)=Y_i=-(1/2\phi)\int_{z0}^{z_i}d\zeta\int_{z0}^{\zeta}E(u)du. \quad (1)$$

where $z_0$ is a point on the axis on the entrance side of the deflection field and $z_i$ is the coordinate of the screen position; $z_0$ and $z_i$ are ordinarily in regions where the field-strength of the deflection field is nearly equal to zero. (The deflection electrodes are assumed to be of infinite extent in the $x$ direction.)

If Eq. (1) is used to determine the deflection, the axial field-strength distribution $E(z)$ must be known. Some electrode configurations exist for which $E(z)$ may be determined mathematically. In most cases, this distribution must be obtained by an experimental method, e.g., one which makes use of an electrolytic tank. Electrode systems of the first type are parallel plates with and without fringing fields, parallel cylinders or wires, and semi-infinite co-planar sheets. A field of the second type is that produced by the commonly employed bent-plate electrode system.

Before going into the discussion of these fields it should be mentioned that an alternative method might be used to determine the deflection $d$ by such fields. This method is based on a formula derived by H. G. Rudenberg[3] which expresses the deflection $d$ in terms of the capacitance per unit width of the deflection system. The derivation of this formula may be based on Eq. (30) of Part I for $Y(z)$ which led to Eq. (1). The slope of the electron path is given by (assuming $y_0'=0$)

$$Y'(z)=-(1/2\phi)\int_{z0}^{z}E(u)du=\tan\theta. \quad (2)$$

---

* Sylvania Electric Products, Inc., 34-10 Linden Place, Flushing, New York.

[1] I. G. Maloff and D. W. Epstein, *Electron Optics in Television* (McGraw-Hill Book Company, Inc., New York, 1938), p. 198.

[2] R. G. E. Hutter, J. App. Phys. 18, 740 (1947).

[3] H. G. Rudenberg, J. App. Phys. 16, 279 (1945).

If the field is essentially different from zero only between the axial coordinates $z_1$ and $z_2$, the angle at $z_2$ for small deflection angles will be given by

$$\theta_2 = -(1/2\phi)\int_{z_1}^{z_2} E(u)du. \tag{3}$$

Rudenberg has shown by an application of Gauss' theorem that Eq. (3) is equivalent to

$$\theta_2 = (1/2\phi)\rho\ \epsilon_0,$$

where $\rho$ is that portion of the charge density on the deflection electrodes which is active in producing the deflection, and $\epsilon_0$ is the dielectric constant of free space. The charge density $\rho$, the capacitance $c$ per unit width of the deflecting electrodes, and the deflection potential $\Delta\phi$, are related by

$$\rho = c\cdot\Delta\phi.$$

Hence,

$$\theta_2 = c\Delta\phi/2\phi\epsilon_0. \tag{4}$$

The deflection at the screen becomes

$$d = (z_s - z_2)\cdot\theta_2 = (z_s - z_2)\cdot(c\Delta\phi/2\phi\epsilon_0). \tag{5}$$

For some electrode configurations it is possible to determine the capacitance per unit width mathematically; in most cases this quantity would have to be measured.

It is apparent that both expressions for the deflection $d$ (Eq. (1) and Eq. (5)) will yield the same result if $z_s$ is greater than $z_2$. The formula given by Eq. (5), however, cannot be used to determine the deflection at the coordinate of the exit side of the electrode configuration because of the contribution of the fringing field on that side of the electrodes to the value of the capacitance $c$. The magnitude of the deflection at that point is important in deciding the question of whether or not an electron beam hits the electrodes. It may be determined by means of Eq. (1).

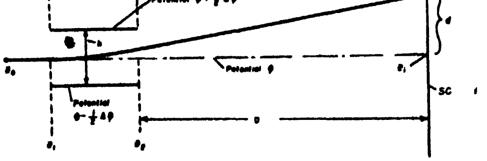

L
Potential $\phi+\frac{1}{2}\Delta\phi$
h
$z_0$
Potential $\phi$
$z_s$
d
SC
Potential $\phi-\frac{1}{2}\Delta\phi$
D
$z_1$
$z_2$

FIG. 1. The parallel-plate deflection field.

## 1. Mathematically Obtained Field Distributions

### a. *Parallel Plates without Fringing Effects*

The field that can most easily be treated mathematically is an idealized one which, however, cannot be produced in practice if the electrodes are of finite length. It is that between two parallel plates which are at the potentials $\phi+\frac{1}{2}\Delta\phi$ and $\phi-\frac{1}{2}\Delta\phi$. The geometry of this field is shown in Fig. 1. The idealization is expressed by making the following assumption regarding the field-strength distribution:

$$E(z) = \left\{\begin{array}{lll} 0 & \text{For} & z<z_1, \\ -\Delta\phi/h & \text{For} & z_1 \leqq z \leqq z_2, \\ 0 & \text{For} & z>z_2. \end{array}\right\} \tag{6}$$

If $E(z)$ of Eq. (6) is substituted in Eq. (1), one obtains the familiar formula

$$d = (\Delta\phi/2\phi)(1/h)L(D+\tfrac{1}{2}L), \tag{7}$$

where $z_2 - z_1 = L$, $z_s - z_2 = D$, $h$ is the distance between the electrodes, $\phi$ is the accelerating potential (potential of the last anode) and $\Delta\phi$ is the potential difference between the electrodes.

Because of the idealization of the field-strength distribution, large errors occur if Eq. (7) is applied to actual parallel-plate deflection systems.

### b. *Parallel Plates with Fringing Effects*

An application of the Schwartz transformation of the theory of functions of a complex variable yields an expression for the equipotential distribution at the edge of semi-infinite parallel plates.[4] It is, therefore, possible to obtain a mathematical expression for the axial field-strength distribution of such an edge field. In order to obtain the axial field-strength distribution of parallel plates of finite length two expressions for a single edge may be used. If the axial coordinates of the entrance and exit edge of the parallel plates are given by $z_1$ and $z_2$, respectively (Fig. 1), the axial field strength may be written as:

$$E(z) = -(\Delta\phi/h)(1/1+r), \tag{8}$$

where $r$ is given in terms of the axial coordinate

---

[4] R. Rothe, F. Ollendorff and K. Pohlhausen, *Theory of Functions* (Technology Press, Massachusetts Institute of Technology, Cambridge, 1942), p. 138.

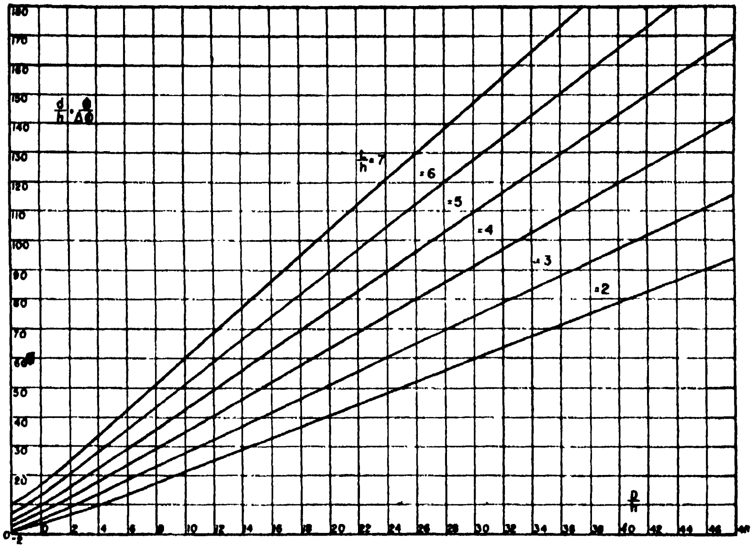


FIG. 2. The deflection by parallel plates (I form).

by:

$$\left.\begin{aligned} z&=z_1+h/2\pi[1+r+\ln r] \\ &\qquad\text{For}\quad -\infty<z\leqq z_1+z_2/2, \\ z&=z_2-(h/2\pi)[1+r+\ln r] \\ &\qquad\text{For}\quad z_1+z_2/2\leqq z<+\infty. \end{aligned}\right\} \quad (9)$$

Measurements have shown that Eqs. (8) and (9) describe the axial field-strength distribution well for plates for which the field strength $E(z)$ is very nearly $-\Delta\phi/h$ at $z=\frac{1}{2}(z_1+z_2)$. This was found to be the case for parallel plates with ratios of $L/h$ larger than two.

The integration indicated by Eq. (1) must be performed in two steps corresponding to the two branches of the axial field-strength distribution. The resultant mathematical expression for the deflection is unwieldy, hence the deflection is better presented in the form of a graph. Any quantity except $h$ can easily be determined from Fig. 2 if the other quantities are known. It is possible to replot the graph and obtain a new one (Fig. 3) which permits the determination of $h$ with ease if the other quantities are known.

### c. *Parallel Cylinders or Wires*

The method of conformal mapping yields in this case a mathematical expression for the axial field distribution.[5] The geometry is shown in Fig. 4. With the notation explained in this figure, the axial field-strength distribution becomes

$$E(z)=\frac{2\Delta\phi}{cq/2}\cdot\frac{1}{\ln\dfrac{1+q}{1-q}}\cdot\frac{1}{1+\left(\dfrac{z}{cq/2}\right)^2}. \quad (10)$$

Substituting Eq. (10) in Eq. (1), the deflection $d$ is given by

$$d=\frac{\Delta\phi}{\phi}\cdot\frac{cq}{2}\cdot\frac{1}{\ln\dfrac{1+q}{1-q}}\left[\frac{z_i}{\dfrac{cq}{2}}\arctan\frac{\dfrac{z_i}{cq/2}-\dfrac{z_0}{cq/2}}{1+\dfrac{z_iz_0}{(cq/2)^2}}+\frac{1}{2}\ln\frac{1+\left(\dfrac{z_0}{cq/2}\right)^2}{1+\left(\dfrac{z_i}{cq/2}\right)^2}\right], \quad (11)$$

which is shown graphically in Fig. 5. $[z_0/(cq/2)]$ was assumed to be equal to $-30$. At this point $E\cong(1/900)E(0)$.

This deflection field was discussed by G. Rudenberg.[3]

### d. *Semi-Infinite Co-planar Sheets*

Another field for which the axial field-strength distribution may be determined accurately by the method of conformal mapping is that produced

---

[5] K. Kupfmüller, *Einführung in die Theoretische Elektrotechnik* (Julius Springer Verlag, Berlin, 1932), p. 71.

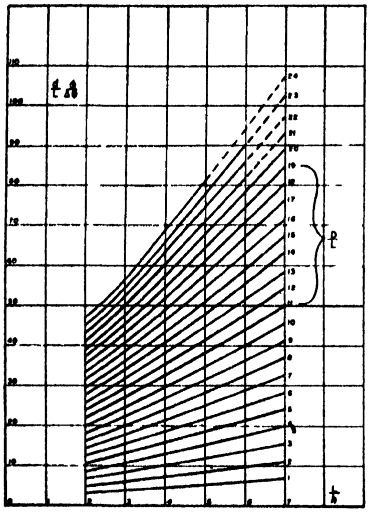


FIG. 3. The deflection by parallel plates (2 form).

by two semi-infinite co-planar sheets.[6] The geometry is shown in Fig. 6. With the use of the notation indicated in this figure, the axial field-strength distribution becomes

$$E(z)=(1/\pi)\Delta\phi(1/a)1\bigg/\left[1+\left(\frac{z}{a}\right)^2\right]^{\frac{1}{2}}. \qquad (12)$$

The deflection is then given by

$$d=\frac{a\,\Delta\phi}{2\,\pi\phi}\left\{\frac{z_i}{a}\ln\frac{z_i/a+[1+(z_i/a)^2]^{\frac{1}{2}}}{z_0/a+[1+(z_0/a)^2]^{\frac{1}{2}}}-\left(\frac{z_i}{a}-\frac{z_0}{a}\right)-\left(\frac{1}{z_i/a+[1+(z_i/a)^2]^{\frac{1}{2}}}-\frac{1}{z_0/a+[1+(z_0/a)^2]^{\frac{1}{2}}}\right)\right\}. \qquad (13)$$

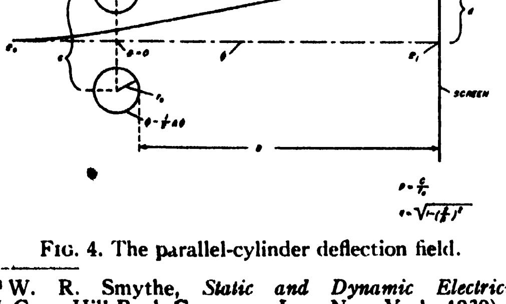


FIG. 4. The parallel-cylinder deflection field.

This expression is shown graphically in Fig. 7. $z_0/a$ was assumed to be equal to $-900$. At this point $E\cong(1/900)E(0)$.

## 2. An Experimental Method to Determine the Field Distribution

In general, it is not possible to determine the axial field distribution mathematically. It may, however, be determined experimentally with sufficient accuracy by means of an electrolytic tank. This instrument is useful in many electron-optical investigations and is available in many electronic laboratories in one form or another.

The theory underlying the use of the tank has been presented many times in the literature[7, 8] and will, therefore, not be repeated here. The following paragraphs will give a brief description of the instrument, developed in the Research Laboratories of Sylvania Electric Products, Inc., and its application to a number of problems of deflection systems.

### a. *The Electrolytic Tank Potential-Plotting Apparatus.*

Figure 8 is a photograph of the entire set-up which consists of two electrolytic tanks ($T_1$, $T_2$), a plotting board ($B$), an instrument rack (I.R.), a plotting mechanism, and an arrangement for positioning the electrodes to be investigated.

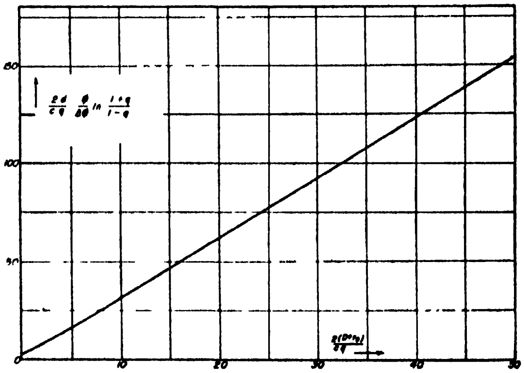


FIG. 5. The deflection by parallel cylinders $\left(z_0/(cq/2)=-30,\ E\left(\frac{z_0}{cq/2}\right)\cong\frac{1}{900}E(0)\right)$.

---

[6] W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company, Inc., New York, 1939), p. 89.

[7] V. K. Zworykin and G. A. Morton, *Television* (John Wiley and Sons, Inc., New York, 1940), p. 73.

[8] V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, and A. W. Vance, *Electron Optics and the Electron Microscope* (John Wiley and Sons, Inc., New York, 1946), p. 389.

These tanks are constructed of lava slabs; the inside dimensions of tank $T_1$ are 30″×30″×60″, and 5″×30″×30″ for tank $T_2$. Ordinary tap water is used as the electrolyte in tank $T_2$, while deionized water is used in tank $T_1$. The plotting mechanism consists of two rails (R.R.) extending the full length of both tanks. A carriage $C_1$ is provided which travels on the rails R.R., and is provided with another pair of rails r.r., at right angles to R.R., upon which is mounted a second carriage $C_2$. The second carriage carries the probe ($P$) and a long beam member ($b$) which holds the stylus ($S$). A close-up view of the tank $T_2$ and the components $C_2$, $b$, and $P$ is shown in the photographs, Fig. 9. The probe is a 5-mil platinum wire which is sealed into one end of a glass tube which is enclosed in a metal tube for shielding purposes. The probe may be brought to any point on the water surfaces of either of the two tanks by turning two hand wheels which operate—through gears—the motion of both carriages. Stylus and probe are rigidly connected and trace out identical curves. Figure 9 also shows a larger sized view of the electrode holders. The electrodes under investigation (in Fig. 8—two bent plates) are supported by adjustable arms (L.L.) which slide on two flat rails ($R_1$, $R_2$). The heads (H.H.) of the support mechanism can be tilted in two directions and the electrodes can be adjusted in the vertical direction. The instrument rack (I.R.) above the tanks holds three units: the potential source ($O$), a single vacuum-tube voltmeter (VTVM), and a unit comprising 5 vacuum-tube voltmeters (5 VTVM's). The potential source is a 500 cycle/sec. resistance-capacitance oscillator which provides a signal of about 50 volts r.m.s. across 500 ohms without distortion of the wave shape. The circuits of all six vacuum-tube voltmeters are alike, the circuit diagram being shown in Fig. 10. The meter reads zero if the signal from the probe is equal to the potential across the calibrated potentiometer.

A block diagram of the complete wiring circuit is shown in Fig. 11. The writing pen is operated electromagnetically from a push button near the hand wheels on the carriage $C_1$.

### b. *Procedure and Measurements*

The large tank, $T_1$, may be used for the study of focusing fields or for the determination of axial field-strength distribution of deflection fields which are not of infinite length in any direction. The shallow tank, $T_2$ is convenient for the investigation of two-dimensional electric fields. The problem of determining the axial field-strength distribution, $E$, for a given electrode system may be solved by measuring two equipotential lines near the axis. The field strength, $E(z)$, is computed as the ratio of the potential difference between the lines to the distance between them at every point $z$. A comparison between mathematical and experimental results is shown in Fig. 12 for the case of two parallel cylinders; the agreement is quite satisfactory. The same tank can also be used for the inverse problem of determining electrode shapes producing a prescribed field-strength distribution $E(z)$. By means of Eq. (20), Part I,[2] two equipotential lines near the axis are computed. The convergence of the series is generally good for small values of $y$. Then the five probes—connected to the unit consisting of five VTVM's—are placed in the tank along the computed equipotential line, and the electrode shapes varied

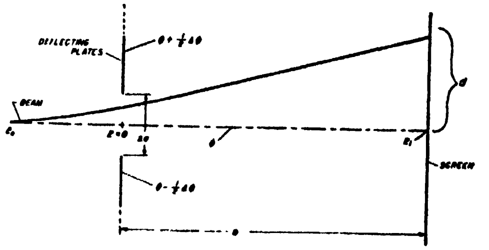


FIG. 6. The semi-infinite co-planar-sheet deflection field.

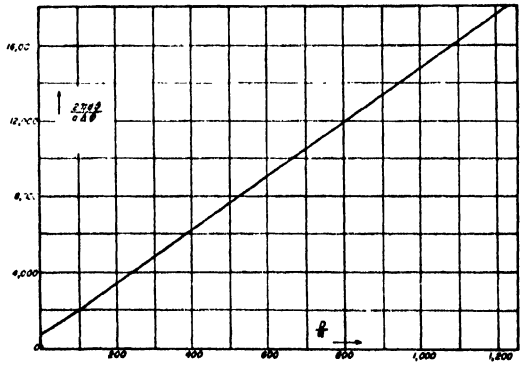


FIG. 7. The deflection by semi-infinite co-planar sheets ($z_0/a=-900$; $\dot{E}(z_0/a)\cong(1/900)E(0)$).

FIG. 8. Potential-plotting tank (Sylvania Electric Products, Inc.).

until all five meters read the potential value of the line simultaneously.

In order to increase the versatility of the tank set-up, an arrangement $T$ tilts the tank $T_2$, thus producing a wedge-shaped volume of electrolyte. In this manner the tank may be used to investigate so-called "Pierce-guns."[9] The five VTVM's are employed to measure the potential distribution along the edge of the electron beam, which is, in this case, represented by an insulator shaped in the form of the desired beam.

### 3. Bent-Plate Deflection Systems

The axial field-strength distributions of a number of bent-plate deflection systems were measured in the manner described in Section 2-b. The geometry of the electrodes and the notation are shown in Fig. 13. The parameters $L_1/h$, $L_2/L_1$, and $\theta$ were chosen to cover a range sufficient to compute the deflection for other values than those for which the graphs are given by interpolation. The integrations according to Eq. (1) were performed by means of a planimeter and the results are plotted in Figs. 14 and 15.

## B. DEFOCUSING AND DISTORTION EFFECTS OF TYPICAL ELECTRIC AND MAGNETIC DEFLECTION FIELDS

Crossed deflection fields, as used in cathode-ray and television tubes, produce undesirable effects of two kinds. One is the beam deformation causing a change in size and shape of the luminous spot on the fluorescent screen of the tube. The other is the distortion of the "pattern." The "pattern" may be the rectangular scanning pattern of the television tube or a Lissajou figure on the screen of a cathode-ray tube in an oscilloscope. Many factors may contribute, and usually do, to the pattern distortion. Only one of these will be investigated here, namely that due to the electric or magnetic field distribution. Discussion of other factors, such as wave-shape of driving circuits, frequency response of the amplifiers, coupling between circuits, or coils of the magnetic deflection yoke, etc., will be omitted.

Idealizations concerning the electron beam and the focusing system mentioned in Part I[2] will be used here.

This method of investigation permits one to distinguish defects caused by the deflection fields from those caused by other components of the device employing such fields. It then becomes possible to attribute a certain observed defect to the proper component.

There are a number of elementary explanations for the mechanism of both types of defects. All of them are qualitative in nature with one exception—that of electron-ray tracing. The electron trajectory through a given potential or field-strength plot is determined by means of numerical or graphical methods of integration. However, unless this method is used with extreme care, the results are not very reliable. In the language of this work, the deflection of an electron at the screen must be determined accurately to distances which are small compared with the quantities here called $\Delta x_i$ and $\Delta y_i$, which in turn are

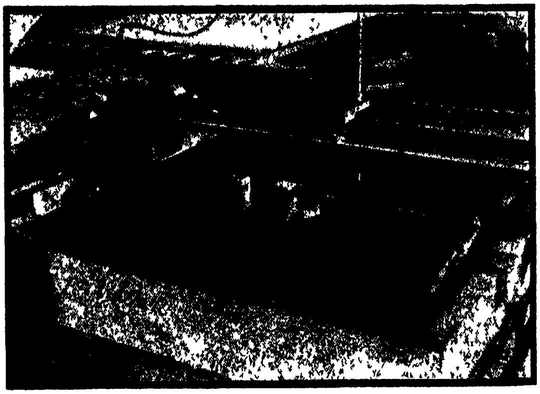


FIG. 9. Partial view of potential-plotting tank.

---

[9] J. R. Pierce, J. App. Phys. 2, 548 (1940).

small compared with the first-order deflection $Y(z_i)$. In order to achieve this accuracy it is necessary to proceed in very small steps in the determination of the trajectory, increasing the amount of numerical work considerably and requiring a very accurate equipotential line or field-strength plot. Furthermore, it is difficult to apply this method to initially deflected electrons and to magnetic deflection fields. In the search for better deflection fields only qualitative hints, at best, were obtained.

The theoretical procedure outlined in Part I, Section 2[2], avoids the difficulties mentioned above. It will now be applied to a number of special deflection fields.

### 1. Special Electric Deflection Fields

The electric deflection fields discussed in this section are those produced by parallel plates (including the fringing fields), parallel cylinders, and bent plates. A comparison of the defocusing and distortion effects of these fields is of interest. In order to make such a comparison it will be necessary that some basis be chosen upon which to compare the defects which are the functions of the accelerating potential $\phi$, the deflection potential $\Delta\phi$, the electrode-to-screen distance, the deflection $d$, and the geometry of the deflection electrodes which determines the axial field-strength distribution $E(z)$. It is possible to formulate various bases of comparison. One way, for instance, requires that the potentials $\phi$ and $\Delta\phi$, the deflection $d$, the distance from the exit side of the electrodes to the screen, and the minimum separation between the electrodes should be the same for otherwise different fields which are being compared. This basis was chosen for the examples discussed below. A better basis might be the following which also requires that $\phi$, $\Delta\phi$, and $d$ should be the same, but which replaces the before-mentioned condition concerning the electrode-to-screen distance and electrode separation by another one. The field-strength distribution will ordinarily be such that $E(z) < E_{\max}/n$ for all axial coordinates $z$ smaller than a value $z_2$ and larger than a value $z_1$, where $n$ may be arbitrarily chosen, so that one could say that the essential part of $E(z)$ lies between the coordinates $z_1$ and $z_2$ (e.g. $n$ a value between 10 and 100). Fields may then be compared for which $\phi$, $\Delta\phi$, $d$, $z_2 - z_1 = L$, and $z_i - z_2$ are the same. Again, keeping $\phi$, $\Delta\phi$, and $d$ equal, fields produced by electrodes occupying a given volume may be compared. The choice of a basis is, therefore, arbitrary and should be made to suit one's particular requirements.

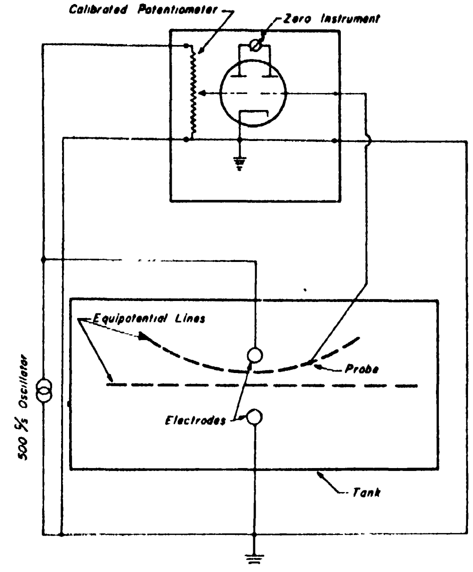


FIG. 11. Block diagram of electrical circuit.

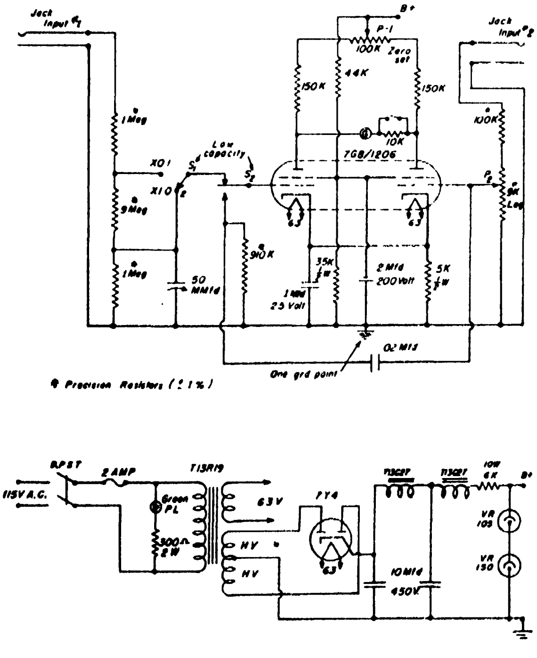


FIG. 10. Circuit diagram of the vacuum-tube voltmeter.

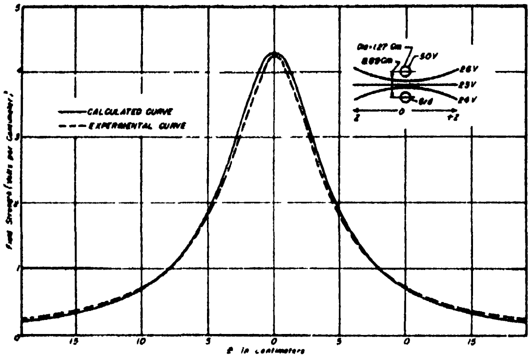


Fig. 12. Comparison between a measured and a computed axial field-strength distribution (parallel-cylinder deflection field).

Figure 16 shows the electrode geometry of three two-dimensional electric deflection fields whose defocusing and distortion effects were computed. This figure shows that the minimum distance between deflecting electrodes was kept nearly the same in all three cases.

Figure 17 shows the spot distortion for the case of a cylindrical beam whose diameter was assumed to be 0.08 inch. The figures show the focusing effect of the deflection field, demonstrating the statement[2] made that electron prisms are always joined to weak cylinder lenses.

Figure 18 shows pattern and spot distortions of a conical beam produced by parallel-plate and parallel-cylinder deflection fields. The deflections were assumed to be those produced by electrodes of dimensions shown in Fig. 16, the horizontal deflection fields preceding the vertical ones. The horizontal deflection potential could be reduced to $\Delta\phi = 140$ volts for a 2.5-inch deflection due to the larger electrode-to-screen distance. The electrode-to-electrode spacings of the horizontal and vertical fields in the axial direction were one inch for the parallel plates and 0.266 inch for the parallel cylinders. The diameter of the conical beam before entering the fields was 0.08 inch, and the apex of the cone was at the center of the fluorescent screen when undeflected. It should be mentioned that the calculations of spot and pattern distortion take into account the defocusing which is caused by a flat screen.

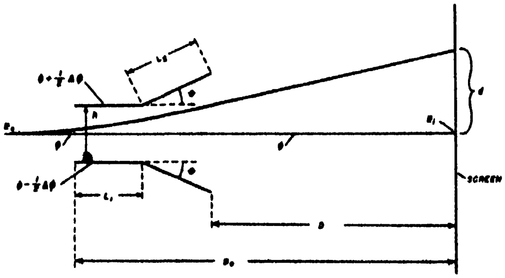


Fig. 13. The bent-plate deflection field.

In the cases investigated so far, it was assumed that perfect alignment of the focusing and deflection systems existed. The distortion figures were therefore mirror-symmetrical with respect to the $x_i$ and $y_i$ axes. In actual tubes, distortions of the spot are observed which do not show this type of symmetry. It was thought that some kind of misalignment of gun and deflection system might be the cause of this. Calculations were made assuming the circumference of the electron beam at $z_0$ to satisfy the equation

$$(x_0-r_b)^2+(y_0-r_b)^2=r_b^2,$$

with $r_b$ again equal to 0.04 inch. The beam was point-focused at the center of the screen. The resulting pattern and spot distortions are shown in Fig. 19 for the case of the parallel-plate field. The distorted spots are shown as little polygons, the corners of which are the points on the circumference of the beam for which the $\Delta x_i$ and $\Delta y_i$ were determined. Actually, the corners will appear rounded off.

### a. *The Parallel-Cylinder Deflection Field*

Most electrically deflected cathode-ray and television tubes use either parallel or bent plates as deflection electrodes. A discussion of such fields with regard to the change of their defocusing effects with the electrode geometry would require many tables. Instead, a textbook-type of example of a deflection field will be discussed here in greater detail. It is that of the parallel cylinders or wires, which was suggested for use where electron transit time and capacitive-shunting effects begin to influence the operation of a cathode-ray tube.[3] It is possible to describe the defocusing effects of parallel-cylinder deflection fields with one set of reduced distortion coefficients for all radii and distances between the wires. These coefficients are given in Table I for various reduced screen distances $[z_i/(cq/2)]$. For

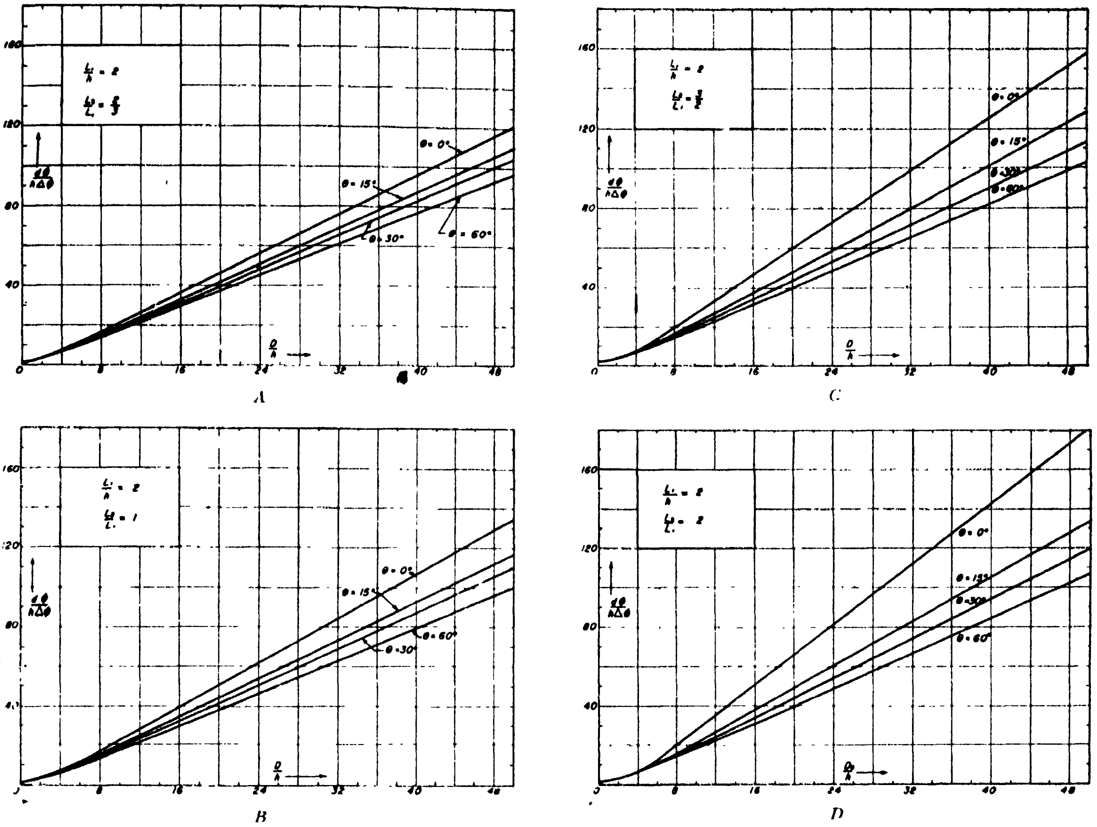


FIG. 14. The deflection by bent plates $A$, $B$, $C$, $D$.

large values of $[z_i/(cq/2)]$ some coefficients are straight lines on linear graph paper, others on logarithmic paper which simplifies the problem of extrapolation.

The notation is the same as was used in Part I.[2] Here $c$ is the center-to-center distance between the cylinders which have the radius $r_0$, and $z_i$ is the distance from the center-to-center plane to the fluorescent screen; $z_0$ is the axial coordinate of a point where the field strength is relatively small compared with that of the maximum at $z=0$. The quantity $(z_0/(cq/2)]$ was chosen as $-30$, i.e.,

$$E_{z0}=(1/901)E_{\max}.$$

The use of the Table to determine a coefficient is simple. As an illustration, one may take

$$\beta_{0000}=\{-31.48\epsilon^3+334.34\epsilon^5-1773.69\epsilon^7\}\cdot c\cdot q,$$

for

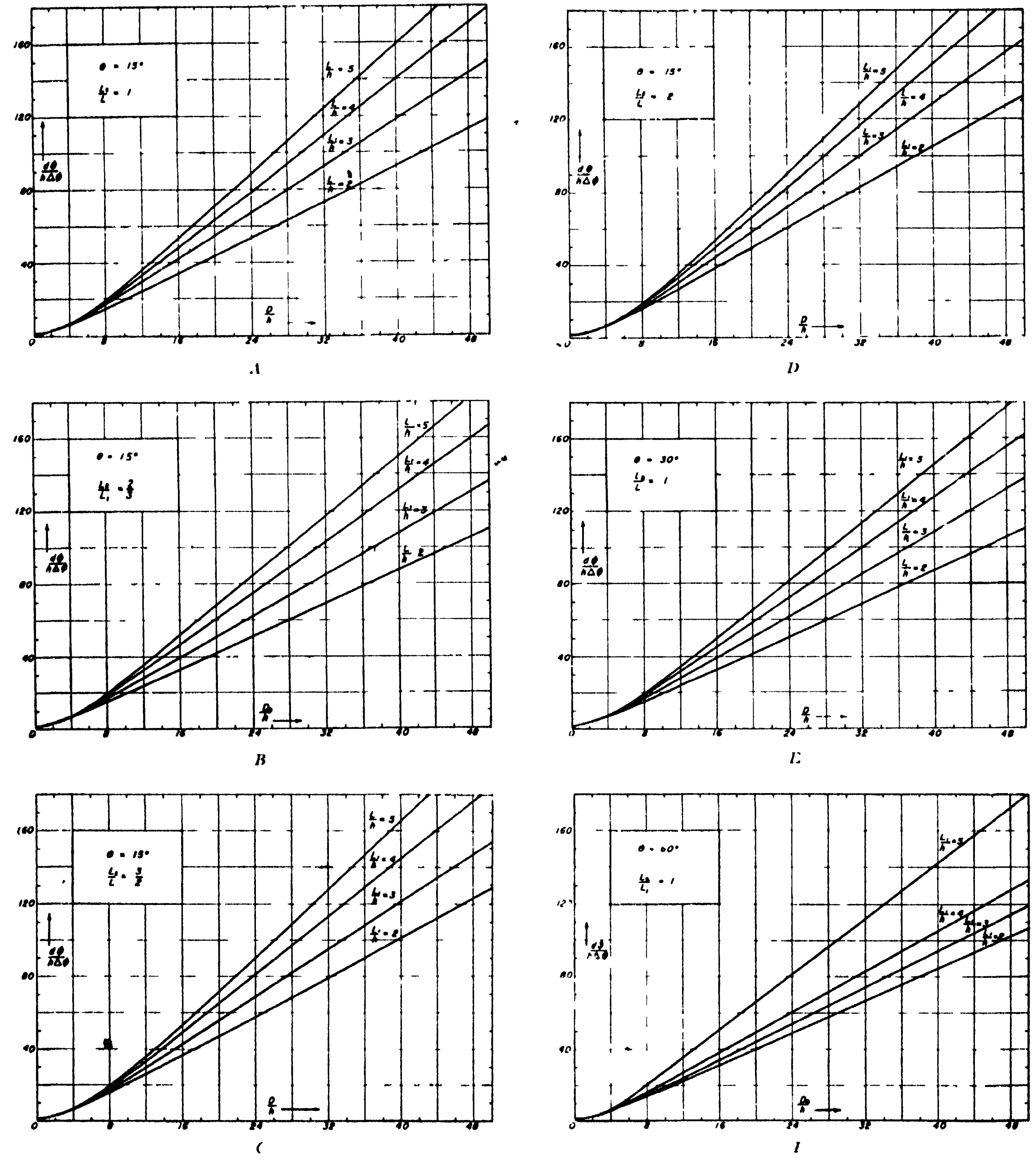
$$2(z_i/cq)=30.$$

The deflection field for the parallel cylinders had the following dimensions: $c=.178''+2\cdot r_0$, $r_0=.598''$, $\Delta\phi=173\,V$, $\phi=2000\,V$, $z_i=10.198''$, and $Y_i=2.5''$. If the minimum distance between the surfaces of the cylinders is kept fixed (0.178″) and the radius is varied in steps from $r_0=.0565''$ to $r_0=23.3''$, the half-width of the field-strength curve ($cq=c[1-(2r_0/c)^2]^{\frac{1}{2}}$) varies between 0.268″ and 4.08″. In order to keep the deflection constant at $Y_i=2.5''$, the deflection potential $\Delta\phi$ will be decreased with increasing value of $r_0$. Keeping the beam dimension the same as in the example shown in Fig. 16, it is found that the aberration coefficients $\alpha_{0010}$, $\beta_{0000}$, $\beta_{0100}$, $\beta_{0001}$, $\beta_{0002}$ decrease rapidly with increasing $r_0$, while $\beta_{0020}$ remains constant over the range of values for $r_0$ given above.

Instead of keeping the distance between the cylinders constant and varying the deflection sensitivity, the latter may be kept fixed and the distance between the cylinders varied, hence

$\Delta\phi$, $\phi$, $z_t$, $V_t$ are kept constant. The larger distance then corresponds to the larger radius. A computation of the distortion coefficients shows that the defocusing effects become less with increasing radius. Similar considerations can be applied to other electrode systems, i.e., if all dimensions except the distances $z_t$ and $V_t$ are multiplied by an arbitrary factor, an improvement of the defocusing effects is obtained. (See paragraph 2.)

It becomes apparent that it is difficult to compare the defocusing effects of different types of electrode systems. It is necessary to introduce assumptions concerning the actual physical size



A D

B E

C F

Fig. 15. The deflection by bent plates *A*, *B*, *C*, *D*, *E*, *F*.

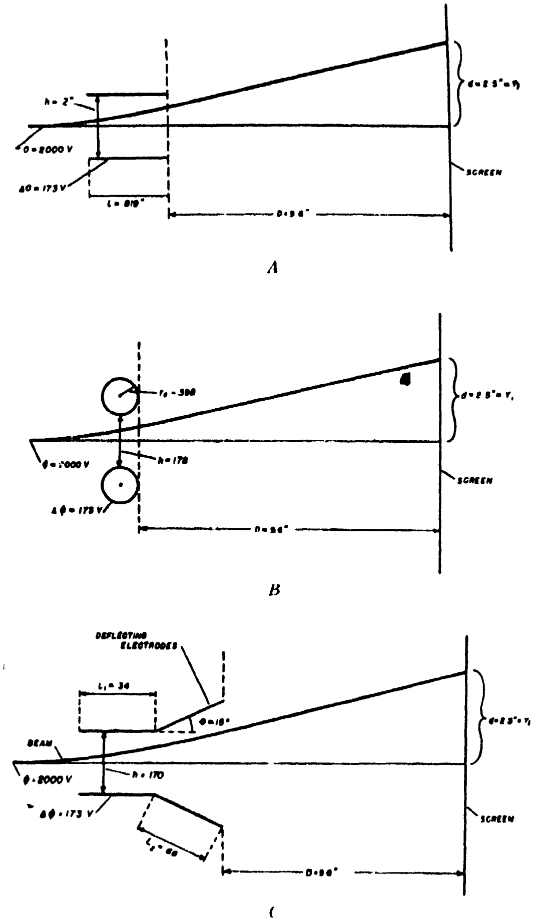


FIG. 16. The geometry of three two-dimensional electric deflection fields.

of the electrodes which are being compared; such assumptions are hard to formulate for electrodes of different shapes.

### b. *A Magnetic Deflection Field*

Pattern and spot distortion were also computed for a special magnetic deflection field. The axial field distributions of the two crossed fields were assumed to be of the form

$$H_x{}^v = -\frac{H_m}{1+(z/b)^2+(y/a)^2},$$

$$H_y{}^h = \frac{H_m}{1+(z/b)^2+(x/a)^2}.$$

The superscripts $v$ and $h$ indicate the vertical and horizontal fields, respectively. The assumption about $H_x{}^v$ means that the field-strength distribution of the $x$ component is a bell-shaped function with a half-width of $2b$ along the $z$ axis and a half-width of $2a$ along the $y$ axis. The half-width in any other direction $y=cz$ in the $(y, z)$ plane will be $[2ab(1+c^2)^{1/2}]/[(a^2+c^2b^2)^{1/2}]$. This half-width is measured in terms of the coordinate $w=(z^2+y^2)^{1/2}=z(1+c^2)^{1/2}$ along the line $y=cz$. A similar interpretation holds for the horizontal deflection field for $H_y{}^h$ in the $(x, z)$ plane. Pole pieces or coil arrangements producing exactly such a distribution were not determined. The magnetic field everywhere else in space is uniquely determined. The functions $V_0$, $V_2$, $H_0$, $H_2$ of the series expressions for the components of $\vec{H}$ (see Part I[2] Eq. (55)) may be determined by expanding $H_x{}^v$ and $H_y{}^h$ in the manner shown for $H_x{}^v$:

$$H_x{}^v = -H_m\left\{\frac{1}{1+(z/b)^2} - \frac{1}{a^2}\frac{1}{[1+(z/b)^2]^2}y^2 - + \cdots\right\}.$$

Hence

$$V_0 = H_m/1+(z/b)^2,$$

$$V_2 = -1/a^2[H_m]/[1+(z/b)^2]^2,$$

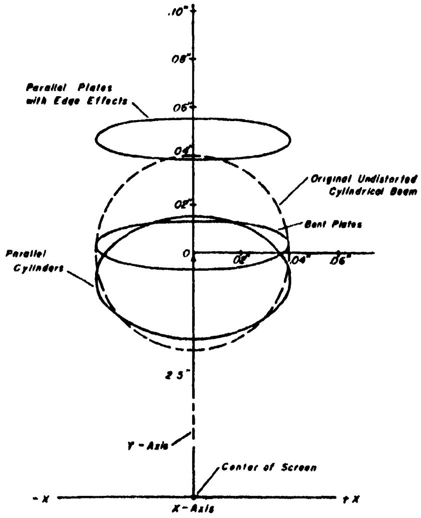


FIG. 17. The spot distortions by three two-dimensional electric deflection fields. (The electrode geometry of these fields is shown in Fig. 1.)

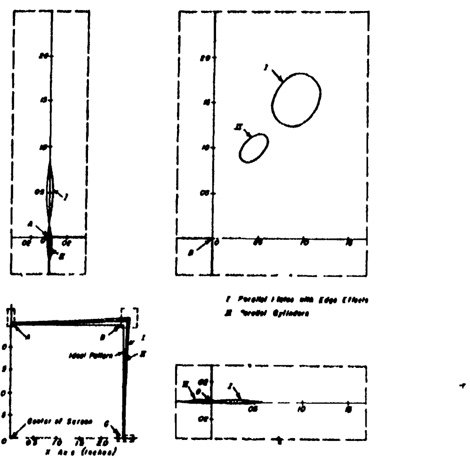


FIG. 18. Pattern and spot distortions by two special crossed electric deflection fields (The electrode geometry and the shape and dimensions of the electron beam are described in the text.)

and similarly for $H_0$, $H_2$

$$H_0 = H_m/1+(z/b)^2,$$

$$H_2 = -1/a^2[H_m]/[1+(z/b)^2]^2.$$

The distortions are shown in Fig. 20 for the following set of parameters: $\phi = 2000$ v, $H_m = 35.4$ gauss, $b = 0.34''$, $z_s = 10.198''$, $z_s/b = 30$, and two values of $a$, namely, $a_1 = 1.36''$, $a_2 = 0.68''$. The maximum deflections on the screen were $d_v = 2.5''$, $d_h = 2.5''$. The electron beam was assumed to be of cylindrical shape with $r_b = 0.04$ inch.

It can be seen from these graphs that the spot distortion increases if the half-width of the field-strength distribution curve, which is in the direction of deflection, is decreased. For very small half-widths, $2a$, the defocusing quantities $\Delta x$, $\Delta y$ become so large that the aberration theory can no longer be applied. A condition similar to that for electrostatic fields Eq. (27) Part I[2] could be formulated for magnetic fields. It is this condition which is violated by allowing $2a$ to become too small.

The graph also shows an elongation of the spot in the direction of the deflection at the corner points of the pattern. This is surprising when compared with the action of one single magnetic field for which an over-focusing is observed in the direction of deflection. A single vertical deflection field over-focuses because the upper edge of the beam passes through weaker portions of the field than does the lower part of the beam. An explanation of the under-focusing of two crossed fields is found when it is discovered that the field-strength function along the line of deflection, $y = x$, $z = 0$, is saddle-shaped, hence causing the outer portions of the beam to pass through stronger fields than the inner parts.

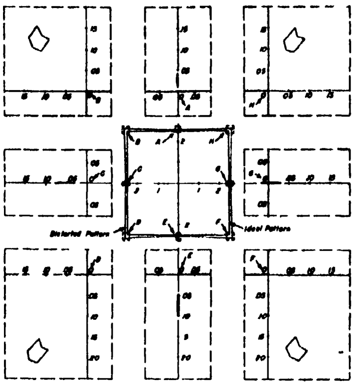


FIG. 19. Pattern and spot distortions by a special parallel-plate field (Electron gun and deflection system were assumed to be misaligned. The electron beam was assumed to point-focus at the center of the screen.)

## 2. *The Problem of Reducing the Defocusing Effects*

At this time only an outline of a method can be given which might lead to deflection fields with less defocusing and distortion.

Each of the coefficients $\alpha_{abcd}$ and $\beta_{abcd}$ determines a characteristic distortion. It was found that $\alpha_{0010}$, $\beta_{0000}$, $\beta_{0020}$ are the important coefficients which determine pattern distortion, while $\beta_{0100}$, $\beta_{0001}$, $\beta_{0002}$ are of primary importance in describing the spot distortion produced by electric deflection fields. It is, therefore, unnecessary to graph every distortion since the magnitude of the coefficients gives the information.

These coefficients may be computed for a number of different fields. A comparison will show which of these is the best so far as they concern defocusing effects. One can then proceed to modify the axial distribution of this field in order to obtain another, having smaller distortion coefficients. The form of the coefficients reveals the "direction" of modification.

The field-strength function obtained in such manner determines the proper configuration of the deflection electrodes or magnets. In case of electric deflection fields, this problem may be solved relatively easily; it is a difficult one in the case of magnetic fields.

In routine design work the method just outlined will not be of much help and is carried out more properly by the development and research groups. A paper written by H. Moss[10] states a theorem which "is of fundamental importance in the design of all-electrostatic cathode ray tubes, since it indicates a general method by which the beam width, and therefore the spot density, can be increased, without an increase in deflection defocusing. Large deflectors and a large neck diameter to accommodate them are used." This theorem reads:

TABLE I. Distortion coefficients of the parallel-cylinder deflection field.

$$\left[\epsilon=\frac{\Delta\phi}{\phi}\,\frac{1}{\ln(1+q/1-q)}\right].$$

| Coefficient | $[z_i/(cq/2)]=5$ | $[z_i/(cq/2)]=10$ | $[z_i/(cq/2)]=20$ | $[z_i/(cq/2)]=30$ | Factor |
|---|---|---|---|---|---|
| $\frac{1}{cq}\cdot\alpha_{0010}$ | 3.24 | 13.14 | 35.12 | 57.93 | $\epsilon^2$ |
| | 6.09 | 6.31 | 6.51 | 6.56 | $\epsilon^4$ |
| $\alpha_{0110}$ | −2.91 | −3.01 | −3.06 | −3.08 | $\epsilon$ |
| | 5.70 | 5.77 | 5.78 | 5.79 | $\epsilon^3$ |
| $\frac{1}{cq}\cdot\alpha_{0011}$ | 15.53 | 31.53 | 61.58 | 92.25 | $\epsilon$ |
| | −13.36 | −27.79 | −56.72 | −85.69 | $\epsilon^3$ |
| $cq\cdot\alpha_{0210}$ | 1.57 | 1.57 | 1.57 | 1.57 | $\epsilon^2$ |
| $\alpha_{0111}$ | −7.82 | −15.70 | −31.42 | −47.13 | $\epsilon^2$ |
| $\frac{1}{cq}\cdot\alpha_{0012}$ | 10.10 | 39.59 | 157.62 | 353.75 | $\epsilon^2$ |
| $\frac{1}{cq}\beta_{0000}$ | −2.80 | −8.10 | −19.63 | −31.48 | $\epsilon$ |
| | 34.61 | 94.16 | 213.71 | 334.34 | $\epsilon^5$ |
| | −282.44 | −580.45 | −1177.06 | −1773.69 | $\epsilon^7$ |
| $\beta_{0100}$ | −5.53 | −15.80 | −37.59 | −59.36 | $\epsilon^2$ |
| | −12.70 | −6.07 | +3.70 | +13.77 | $\epsilon^4$ |
| | −460.13 | −936.38 | −1888.88 | −2841.38 | $\epsilon^6$ |
| $\frac{1}{cq}\cdot\beta_{0001}$ | 38.78 | 122.37 | 459.89 | 1013.95 | $\epsilon^2$ |
| | 50.27 | 88.68 | 101.69 | 10.93 | $\epsilon^4$ |
| | 1122.94 | 4614.19 | 18,752.00 | 42,317.00 | $\epsilon^6$ |
| $cq\cdot\beta_{0200}$ | −.0038 | +.0369 | +.01242 | +.00888 | $\epsilon$ |
| | −26.81 | −51.78 | −101.44 | −151.16 | $\epsilon^3$ |
| | −251.85 | −508.35 | −1021.35 | −1534.35 | $\epsilon^5$ |
| $\frac{1}{cq}\cdot\beta_{0020}$ | 7.77 | 15.77 | 30.79 | 46.13 | $\epsilon$ |
| | −6.68 | −13.89 | −28.36 | −42.85 | $\epsilon^3$ |
| $\frac{1}{cq}\beta_{0002}$ | −.055 | +2.005 | 3.705 | 5.36 | $\epsilon$ |
| | −157.13 | −1320.73 | −10,407.92 | −34,747.84 | $\epsilon^3$ |
| | −1552.28 | −12,556.28 | −101,329 | −343,229 | $\epsilon^5$ |
| $\beta_{0101}$ | 2.87 | 2.93 | 2.74 | 2.66 | $\epsilon$ |
| | 133.93 | 527.09 | 2061.97 | 4599.98 | $\epsilon^3$ |
| | 1246.80 | 5047.80 | 20,329.8 | 45,886.8 | $\epsilon^5$ |
| $(cq)^2\beta_{0300}$ | −10.40 | −20.80 | −41.60 | −62.40 | $\epsilon^2$ |
| | −46.00 | −94.00 | −191.00 | −288.00 | $\epsilon^4$ |
| $\beta_{0120}$ | −3.91 | −7.85 | −15.71 | −23.57 | $\epsilon^2$ |
| $\beta_{0102}$ | −183.21 | −1536.11 | −12,432.3 | −41,968.2 | $\epsilon^2$ |
| | −884.55 | −7094.10 | −57,388 | −194,533 | $\epsilon^4$ |
| $\frac{1}{cq}\beta_{0021}$ | 10.10 | 39.59 | 157.62 | 353.75 | $\epsilon^2$ |
| $cq\cdot\beta_{0201}$ | 76.45 | 310.45 | 1246.37 | 2806.37 | $\epsilon^2$ |
| | 348.00 | 1413.00 | 5733.00 | 12,963 | $\epsilon^4$ |
| $\frac{1}{cq}\beta_{0003}$ | 140.82 | 2520.36 | 41,322 | 210,054 | $\epsilon^2$ |
| | 755.74 | 11,897.48 | 191,590 | 973,326 | $\epsilon^4$ |

"If the beam width and scale of the whole deflectors (including their spacing) are multiplied by $k$, then the increase of spot size on deflection through a constant angle is unchanged, provided the distance between the screen and the center of deflection is also unchanged."

The proof of this theorem was based on theories of scale, energy, dimensional homogeneity, and experiment. A more rigorous proof showing at the same time some limitations of the scaling process can be based on the theory developed here.

Consider a point-focused (conical) electron beam. The defocusing effects will be given to a first approximation by

$$\Delta y_i=\beta_{0000}+\beta_{0001}y_{iu}'.$$

If *all* dimensions are now increased by a factor $k$ (keeping all potentials constant), the deflection $d$ will very nearly increase by this factor $k$ as long as the electrode-to-screen distance is sufficiently large. The coefficients $\beta_{0000}$ and $\beta_{0001}$ will also increase by the same factor and, since $y_{iu}'$ remains unchanged, $\Delta y_i$ will increase to $k\cdot\Delta y_i$. If the screen is now brought back to its original

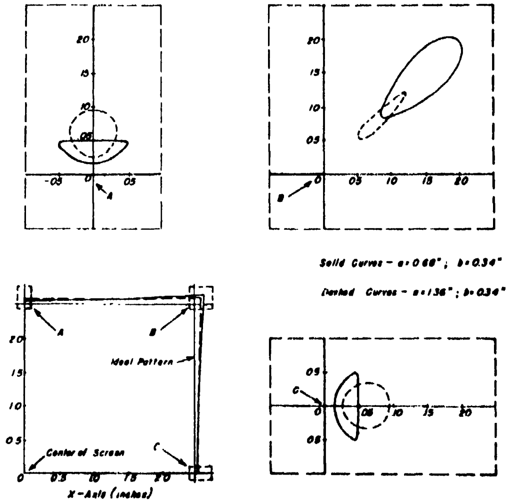


FIG. 20. Pattern and spot distortions by a special pair of crossed magnetic fields. (The fields are given by: $H_x^v=-\frac{H_m}{1+(z/b)^2+(y/b)^2}$, $H_y^h=\frac{H_m}{1+(z/b)^2+(x/a)^2}$)

---

[10] H. Moss, J. Tel. Soc. 4, 206 (1946).

position and the electron beam is refocused to a point at the center of the screen, the deflection will again be equal to $d$. The angle $y_{iu}'$, however, will be $k$ times its original value (for small angles). An inspection of the coefficients $\beta_{0000}$ and $\beta_{0001}$ shows that $\beta_{0000}$ is a linear function of the electrode-to-screen distance, while $\beta_{0001}$ is a quadratic function of this distance (for large electrode-to-screen distances). Decreasing the electrode-to-screen distance restores, therefore, $\beta_{0000}$ to its original value. Since $\beta_{0001}$ increased to $k$ times its original value, and since $\beta_{0001}$ is a quadratic function of the electrode-to-screen distance, $\beta_{0001}\cdot y_{iu}'$ will reduce to its original value when this distance is decreased to its original value. Hence $\Delta y_i$ has the original value.

## ACKNOWLEDGMENT

I would like to thank Mr. L. H. McKee for his assistance in the experimental work, and Miss Gloria Glaston and Mrs. Shirley Harrison for their help with the computations. I also would like to thank Dr. R. M. Bowie, Manager of Research of the Sylvania Electric Products Research Laboratory, for helpful discussions and constant encouragement in our work.

---

# A 70-Mev Synchrotron

F. R. Elder, A. M. Gurewitsch, R. V. Langmuir, and H. C. Pollock
*Research Laboratory, General Electric Company, Schenectady, New York*

(Received April 22, 1947)

A synchrotron for the production of 70-Mev x-rays has been built and tested. The general mode of operation, and various features of its design and construction, are discussed. Some preliminary information on the characteristics of the electron and x-ray beams is included. This work has been supported by the Office of Naval Research.

## I. INTRODUCTION

Soon after the synchrotron principle was proposed by Veksler[1, 2] and McMillan[3] for the acceleration of electrons to high energies, the construction of a 70-million-volt machine, using betatron injection,[4] was begun. The completed synchrotron (Fig. 1) has been in operation at the 70-Mev level for several months. The results obtained with it confirm and extend the experimental results of Goward and Barnes[5] with an 8-Mev machine. The synchrotron provides a satisfactory means of producing high energy electrons and x-rays. By varying the important parameters of the present machine, information is being acquired to facilitate the design and construction of larger units.

The general theories of both the betatron and the synchrotron have been described in earlier publications.[6–10] This particular machine may be regarded as a betatron during the initial portion of each cycle of the magnet. The electrons, after their injection, are accelerated by the changing magnetic flux carried by iron laminations within their orbit. Later, these central flux bars of the magnet begin to saturate and the acceleration is supplied instead by the electric field at the gap of a radiofrequency resonator.

The magnet is driven at 60 cycles from an induction regulator. The magnetizing coils are resonated with a 3000-kva capacitor bank. As the magnetic field begins to increase, electrons are injected into the vacuum tube from a betatron-type gun pulsed with a voltage of the order of 40 kv. The gun filament is pulsed negative for 2 to 3 microseconds in each cycle, the timing of

---

[1] V. I. Veksler, Comptes rendus U.S.S.R. **44** [9], 365 (1944).

[2] V. I. Veksler, J. Phys. U.S.S.R. **9**, 153 (1945).

[3] E. M. McMillan, Phys. Rev. **68**, 143 (1945).

[4] H. C. Pollock, Phys. Rev. **69**, 125 (1946). H. C. Pollock, R. V. Langmuir, F. R. Elder, J. P. Blewett, A. M. Gurewitsch, and R. L. Watters, Phys. Rev. **70**, 798 (1946).

[5] F. K. Goward and D. E. Barnes, Nature **158**, 413 (1946).

[6] D. W. Kerst, Phys. Rev. **60**, 47 (1941).

[7] D. W. Kerst and R. Serber, Phys. Rev. **60**, 53 (1941).

[8] D. Bohm and L. Foldy, Phys. Rev. **70**, 249 (1946).

[9] N. H. Frank, Phys. Rev. **70**, 177 (1946).

[10] D. M. Dennison and T. H. Berlin, Phys. Rev. **70**, 58 (1946).

the pulse being adjustable from $t=0$, when $B_0=0$, to $t=20$ microseconds. This timing is accurately controlled by a multivibrator circuit which is initially triggered from a permalloy peaking strip mounted across the magnet gap. Since the central flux bars do not possess uniform permeability during the betatron acceleration, the betatron orbit is slowly shrinking as $t$ increases. It matches the theoretical synchrotron orbit, which is slowly expanding, at about 200 microseconds. At this time, when the electron energy is about 2 Mev, and the electron velocity 97 percent that of light, another multivibrator circuit turns on a radiofrequency oscillator which excites a 163-megacycle cavity, forming a part of the vacuum envelope. The electrons become bunched in orbital distribution by the electric field at the gap of the resonator, and on successive revolutions they are held in synchronism with the r-f field with a stability which is characteristic of the synchrotron. Following the turn-off of the cavity resonator, the electron energy will remain essentially fixed. If the magnetic guide field has not reached its peak, the electron beam contracts to strike a tungsten-wire target placed within the tube inside the synchrotron orbit. If the field has passed its peak when the resonator is turned off, the beam expands to hit the back of the injection gun and accordingly the x-ray beam comes from the gun.

When the magnet is fully excited, the peak value of $B$, at the orbit radius of 29.3 cm, is 8100 gauss, which permits a peak electron energy of 70 Mev. At present the x-ray yield in the center of the beam at 1 meter from the tungsten target is 50 roentgens per minute as measured by a Victoreen thimble in a $\frac{1}{8}''$ thick lead shield. This output undoubtedly will improve as the many variables of the equipment are adjusted to their optimum settings.

## II. THE MAGNET

The 70-Mev synchrotron magnet (Fig. 2) resembles in general design and construction the betatron magnets which have been built in this laboratory.[11, 12] The total weight of the magnet and coils is approximately 8 tons.

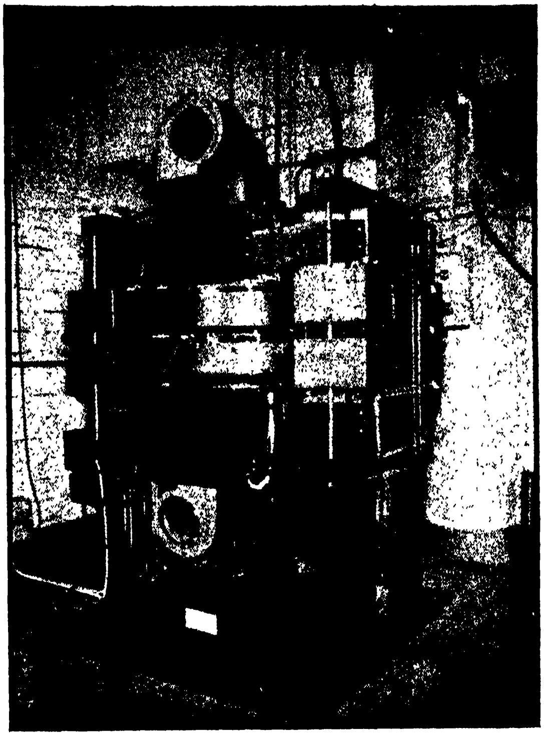


FIG. 1. Photograph of 70-Mev synchrotron.

The shape of the pole face was determined by using a half-scale model made of solid steel. The magnetic field intensity in the plane midway between pole faces varies inversely as the $\frac{3}{4}$ power of the radius. The width of the pole face near the orbit is 4″, and the gap at the 23″ diameter orbit is $2\frac{3}{4}''$.

The pole pieces were assembled from approximately 4° sectors made of enamelled punchings of dynamo-grade silicon steel 0.014″ thick. Each sector was impregnated with permafil and baked while held in a special clamp. The individual punchings, which were properly profiled, varied from 1″ to 12″ wide in steps of 1″. To avoid excessive eddy currents in the pole pieces, a total of 3 sheets of 0.005″ paper was equally spaced between bunches of laminations during the stacking of each sector. A 2″ wide copper sheet, which extended about $\frac{1}{4}''$ beyond the inner end of the iron stack, was placed in the middle of each sector. To assemble accurately pole pieces of 27″ O.D., the sectors were placed on a ground steel plate of this diameter and clamped together with a Textolite ring. The ring, fabricated from

---

[11] W. F. Westendorp, Phys. Rev. 71, 271 (1947).

[12] W. F. Westendorp and E. E. Charlton, J. App. Phys. 16, 581 (1945).

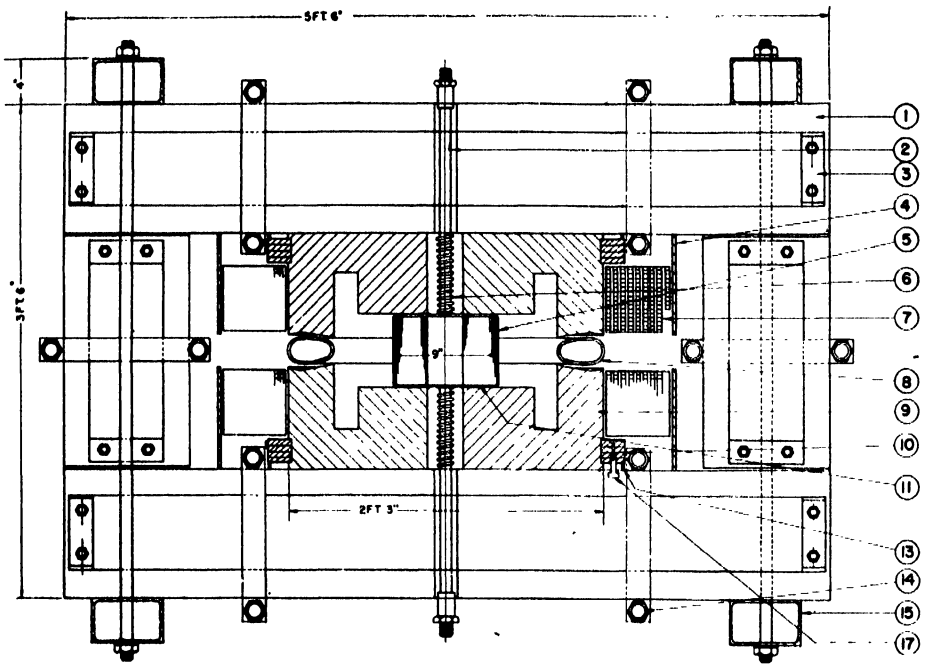


FIG. 2A. Synchrotron magnet—front view.

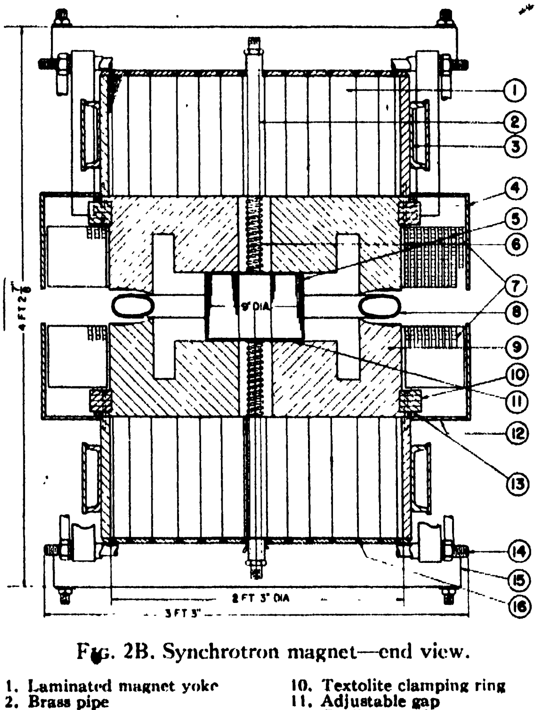


FIG. 2B. Synchrotron magnet—end view.

1. Laminated magnet yoke
2. Brass pipe
3. Clamping beam
4. Herkolite coil housing
5. Flux bars
6. Cooling spiral
7. Magnetizing coils
8. Vacuum tube
9. Pole piece
10. Textolite clamping ring
11. Adjustable gap
12. Cooling opening
13. Exciting coil
14. Clamping bolt
15. Clamping beam
16. Cooling pipe
17. Hanger

6 semicircular pieces, was fastened together with brass pins after assembly in the rectangular groove at the outer edge of the sectors. The joints in the ring were staggered by 120°.

In the central 3″ diameter hole of the pole piece was inserted a spiral cooling coil of copper together with a $\frac{3}{4}''$ brass pipe inside the coil. The space between the brass pipe and sectors was filled with soft solder to provide heat transfer from the copper fins in the sectors to the spiral coil. The brass pipe, long enough to extend through a 2″×2″ hole through the yokes, then served as a support for the central part of the pole piece.

For the yokes and legs of the magnet, enamelled transformer-grade silicon steel .014″ thick was used. For the purpose of cooling, pairs of .030″ copper sheets were distributed at appropriate intervals during the stacking of the yokes and legs. Water flowing through copper tubes, soldered in a V groove formed by bending the outside edges of the copper sheets, serves to remove the heat produced in the iron. The cooling fins also provide a means of attaching the pole pieces to the yokes. This was accomplished by punching T-shaped notches in the sheets into which beryllium-copper hangers were inserted during stacking. By spacing the fins suitably, and by varying the distance of the notches from the center line of each yoke, the hangers were arranged to lie at equal intervals on the circumference of a circle. The circular pole pieces were then attached securely to the yokes with thirty-two brass bolts through each of the Textolite clamping rings, previously assembled in a rectangular groove in the pole piece.

The flux bars of 0.004″ enamelled strip steel consist of twenty-four bundles of strips, $5\frac{15}{16}''$ long. Each bundle contains approximately 120 strips of average width 2″ and is supported in a Textolite holder. Two spacing disks, of about 0.128″ thickness, position the holder between the pole pieces. A stream of air introduced through one of the brass pipes cools the flux bars.

### III. THE COILS

The magnetizing coils, placed around the pole pieces, carry current to produce both flux through the orbit and guiding field at the orbit. Approximately 35,000-ampere turns r.m.s. are needed for 70-Mev operation. The apparent power is furnished at 22,900-volts r.m.s. from a capacitor bank through a high voltage cable to the coils. The two coils are connected in series (Fig. 3) and each contains 130 turns of cable. The cable which consists of 12 strands of 0.050″×0.150″ Formex-covered copper wire is taped together in a bundle while being wound. Glass cloth spacers, treated with permafil, serve to provide insulation between coil layers and to form ducts for forced air cooling. After the coils were wound, they were impregnated in vacuum with permafil and then baked to produce a rigid structure. The coils are supported and held in place inside Herkolite cylinders by insulating blocks bolted to the cylinders. The cylinders, which are fastened to the yoke structure by four brackets, carry through the coils the air flow from two ½-hp blowers.

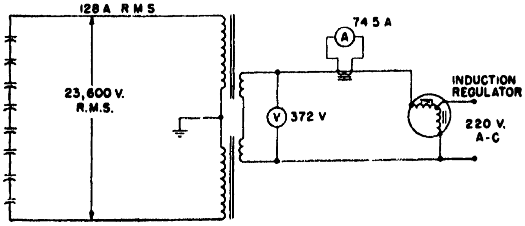


FIG. 3A. Main power circuit.

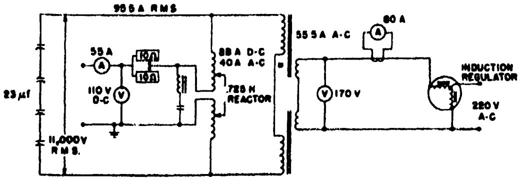


FIG. 3B. Main power circuit with d.c. bias.

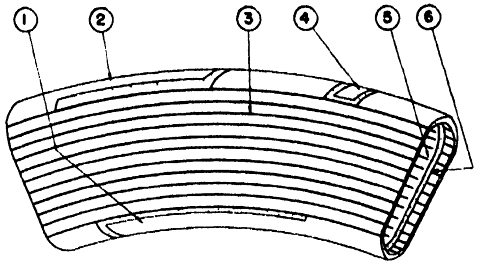


FIG. 4. Pictorial view of cavity.

1. Tuning strip
2. Coupling strip
3. Slots in silver to reduce eddy currents
4. Insulated patch for voltage monitor
5. Connection between adjacent strips
6. Accelerating gap

The exciting coils are connected across an induction regulator which supplies 74.5 amperes at 372 volts for 70-Mev operation of the synchrotron. Each coil has two turns of No. 2 insulated code wire wrapped around the pole piece between the Textolite clamping ring and the yoke, as is shown in Fig. 2.

The saturating coils, two of which are mounted on each leg of the magnet, are required for starting the machine. The inductance of the magnet varies during each cycle as the flux bars saturate. When the machine is not running, its inductance and the capacitor bank are in tune at approximately 40 cycles. Accordingly, in starting, the exciting coils draw an excessive current from the 60-cycle line. Therefore, the saturating coils are supplied with 60 amperes of d.c. which, by partially saturating the magnetic structure, tunes the circuit to 60 cycles for low power operation. As the excitation is increased, the d.c. is reduced by an automatic phase-sensitive control circuit which keeps the magnet-driving current and voltage nearly in phase. Finally, when the magnet is fully excited, no d.c. remains in the saturating coils. Each of the four coils has 46 turns, wound in a single layer on an insulating frame. The individual coils are connected in series so that the a.c. voltages which are induced in them balance out and do not appear across the d.c. supply.

## IV. THE CAPACITOR BANK

The 3000-kva capacitor bank, which supplies the apparent power for the magnetizing coils, consists of 121 units arranged with 8 banks in series. Seven banks have 15 units in parallel, and one bank 16 units in parallel. Each Pyranol-filled unit is rated at about 7 micro farads, 3 kv r.m.s., and 24 kva. The 8 insulated metal racks holding the capacitors are contained within a sheet-steel housing through which air is circulated by a low pressure blower. The power loss in the capacitor cabinet is ⅓ of 1 percent of the voltamperes, or approximately 10 kw.

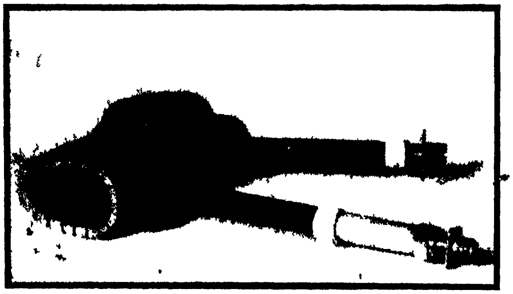


FIG. 5A. Photograph of r-f resonator

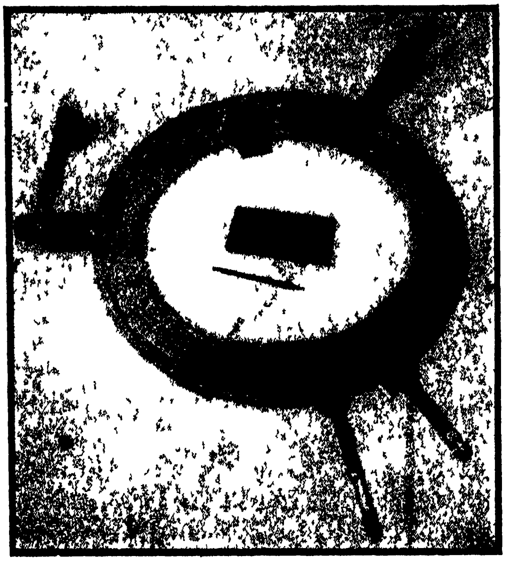


FIG. 5B. Photograph of resonator installed in vacuum tube

## V. VACUUM TUBE AND R-F CAVITY

The vacuum tube was molded by the Corning Glass Works of 774 Pyrex glass, ¼" thick. It is of toroidal shape with an outside diameter of 27", an inside diameter of 19", and with a nearly elliptical cross section; the major axis is 4" and the minor axis is 2¼". Two side arms are provided 120° apart for gun and target, the pumping lead being on the target arm. From the tube, a 57° sector was removed by making radial cuts with an emery wheel. This sector was replaced by one of similar shape, but of 707-DG glass, the loss tangent of which at 163 megacycles is 0.0012, a value low enough to make this glass suitable for the radiofrequency resonator. The 707-DG glass sector, cast in the same mold as the original tube, had its ends ground so that it fitted tightly in the place of the sector which was cut out. The resonator (Fig. 4) was prepared in the following way. The low loss glass was coated appropriately with DuPont silver paint which was baked onto the glass at 500°C. In an electrolytic plating bath, the thickness of the silver coating was increased to 0.001". A tungsten wheel, mounted on a rotating arm and connected to a current transformer, was then used as a tool for removing silver from the glass so as to subdivide the silver coating into narrow strips. This reduces the eddy currents produced in the coating by the a.c magnetic field.

The complete cavity (Fig. 5) may be regarded as a ¼-wave concentric resonator which has been fore-shortened physically by the introduction of glass between inner and outer conductors. Voltage is developed across a ⅛" wide gap inside the sector close to one end. The r-f excitation from the oscillator is fed from a coaxial 50-ohm line. Coupling is achieved by having the central conductor of the cable make proper contact with a strip of the outer silver coating of the cavity. The coupling can be varied to secure a good match by moving the point at which the center conductor of the cable makes contact with the silver strip. This strip is only connected to the remainder of the silver coating at the extreme end of the cavity, away from the gap. It is not necessary to bring any metal leads through the glass. For measuring the voltage across the gap, a calibrated probe and crystal is supported outside the glass near the gap. All the electrical connections are made by silver strips supported on Textolite blocks which are taped to the resonator by permafil-treated glass cloth.

Before the silvered cavity was inserted into the main tube, the inner surface of the tube was chemically coated in a uniform manner to obtain a resistance of 200 to 1000 ohms per square cm. When the cavity is permanently positioned, 0.002″ gaskets are used at the joints. Molded sleeves of butyl rubber hold cavity, gun, and target in position. These sleeves when painted with Glyptal insure vacuum-tight joints. A glass oil diffusion pump of 20 liters per second capacity, backed up by a suitable fore-pump, keeps the tube pressure below $10^{-5}$ mm of mercury. This pressure, measured with an ionization gauge, is satisfactory for regular operation.

The cavity requires 250 watts to produce 1000 volts peak across the gap. As the duty cycle is normally 25 percent, considerably less average power is used. Operation of the cavity shows that the voltage can be developed across the gap without excessive heating of the dielectric or the rubber sleeves. The $Q$ of the cavity has been measured as 375. It is possible to change the resonant frequency 10 percent by moving the position of a piece of foil attached to the tuning strip outside of the cavity. In practice, this tuning is done only to adjust a new cavity to the desired frequency. The power is supplied to the cavity by an 832-A oscillator driving an 829-B grid-modulated power amplifier.

## VI. ORBIT LOCATION

The location of the betatron orbit is of interest only during the time that the magnetic field is low. The orbit is located by methods which are similar to those already published.[12] The voltage induced in a small coil placed at the desired orbit position is balanced against a fraction of the voltage from a single turn of wire of 23″ diameter on the pole face. A null-voltage indication on an oscilloscope enables the orbit position to be calculated. This null indication is observed on a sweep which is triggered when the magnetic field passes through zero. By varying the fraction of the single-turn voltage which is used, the null indication can be obtained at various times following $B=0$. Thus the location of the orbit as a function of time can be found. Adjustment of the betatron orbit is readily carried out by changing the paper shims between the yokes and legs of the magnet. Since the gap of approximately 0.25″ between the central flux bars and pole pieces varies more rapidly, relatively, than the main gap of 2.75″, the flux within the orbit can be altered by a large amount without appreciably affecting the flux density at the orbit. In this way, the betatron condition can be fulfilled in order to match the desired orbit diameter.

The orbit shrinks considerably as the flux bars begin to saturate. The contraction in orbit radius is about 1½″ in 350 microseconds, with most of the contraction occurring near the end of this period. This contraction does not prevent a satisfactory transition to the synchrotron condition. Indeed, the exact location of the betatron orbit in the center of the glass vacuum tube does not appear important. The orbit has been shifted over a wide region inside the tube with little change on the x-ray output. The synchrotron orbit is defined by the energy of the electrons and the frequency of the r-f oscillator.

## VII. POWER CIRCUIT

A diagram of the main power circuit of the machine is shown in Fig. 3A. Power is supplied from a 220 volt 60-cycle a.c. line through a motor-controlled induction regulator to the exciting coils of the magnet. The $Q$ of the resonant circuit is approximately 113. Variations in line frequency could have a relatively large effect on operation, but trouble from this source has rarely been observed with this installation. If the capacitor bank is cold, and hence the capacity high, the machine will be in resonance at a slightly higher power level because of the nonlinear inductance of the magnet. After ten minutes of running, the temperature of the capacitor bank has reached equilibrium. The power required to energize the machine fully is 26.5 kw.

The machine has also been run with d.c. bias on the magnetizing coils as is shown in Fig. 3B. The large increase in electron energy which is obtainable by biasing a betatron[13] of given

---

[13] W. F. Westendorp, J. App. Phys. 16, 657 (1945).

Fig. 6. Oscillogram showing operation of the machine; the trace is cut off every 100 microseconds. (1) marks the time of electron injection. At (2) the r-f voltage is turned on and at (3) it is turned off. (4) marks a signal from a photo-multiplier tube placed in the x-ray beam. For this trace the energy was about 20 Mev.

weight cannot be realized with a synchrotron. However with a biased synchrotron, for a given energy output, the voltage across the magnetizing coils may be reduced by nearly 50 percent and the effective value of the current by about 15 percent. Because of the lowered a.c. component of flux, the iron losses decrease to about $\frac{1}{3}$, which reduces the cooling requirement of the magnet. The a.c. power input is also lowered. Theoretically, the capacitor bank is reduced to about 30 percent of that required for full a.c. operation. The full saving in capacitor kva cannot be realized since it is necessary to compensate the inductance of the reactor required to introduce the d.c. bias current into the magnetizing coils. The net saving in practice may thus amount to about 50 percent. With the present machine, the cost of the reactor was comparable to that of the added capacitor kva required for full a.c. operation. For a higher energy machine, some saving in cost might be obtained by using bias.

With d.c. bias, it was somewhat more difficult to start up and get the machine on the line. The operation was also rather unstable. A large part of the instability was due to the regulation characteristics of the input circuit coupled with the non-linearity of the reactance of the magnet. Normally, the input lines have a small inductive reactance. When the magnetizing circuit is near resonance at the line frequency, the operation will be stable provided the power factor of the circuit is slightly lagging. A small change in the d.c. bias current can cause the power factor to swing from lagging to leading or *vice versa.* If the power factor becomes leading, then the voltage at the machine terminals rises because of the inductive reactance of the input lines. A rise of voltage on the exciting coils causes a higher effective inductance of the magnet, and hence a tendency for the power factor to become still more leading, since the capacitor bank has a fixed value. Obviously, under such conditions, the operation will be unstable. One way in which this difficulty can be overcome is by means of a series capacitor in the input circuit. This was tried and found to be reasonably effective. We are confident that stability can be obtained when it is desirable to use a biased magnet.

With full a.c. operation, there is stability. An increase of voltage at the input terminals decreases the effective inductance of the magnet, there by making the power factor more lagging, which tends to lower the input voltage.

### VIII. OPERATION

The operation of the machine is quite simple. The sequence of events in each cycle is illustrated by Fig. 6. The final energy of the electrons is easily changed by a single knob controlling the time that the r-f cavity stays on. Figure 7 shows that the time at which the r-f is turned on is not very critical. The time at which the electrons are injected into the tube is critical as is shown by Fig. 8. Under good conditions, when the line frequency is stable, the injection timing multivibrator is adjusted perhaps once or twice a minute. It is easy to keep the intensity of the x-rays constant to within about 10 percent for long periods of time.

The transiton from betatron operation to synchrotron operation seems to be rather efficient.

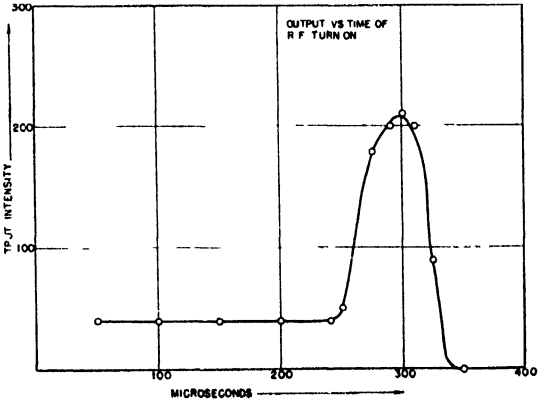


Fig. 7. Output *vs.* time r-f is turned on.

This is shown by operating the machine at a low intensity and observing the output by means of a Geiger counter. The pulses from the counter are displayed on an oscilloscope with a 60-cycle sweep. If the r-f oscillator is not turned on, a count is registered almost every cycle a few hundred microseconds after injection, at the sweep position where 3.5-Mev x-rays are expected. This is of course the betatron beam collapsing to the target as the central flux bars saturate. If the r-f is turned on, no betatron counts are observed, but higher energy quanta produce counts following r-f turn off. This indicates that a large fraction of the electrons which previously had collapsed to the target are bunched in the synchrotron orbit and are accelerated to high energy.

The rate of rise of the r-f envelope was varied from 2 to 20 microseconds and does not seem to affect the efficiency of the transition. Increasing the peak r-f voltage on the accelerating gap beyond twice the minimum voltage needed per turn does not increase the output appreciably (see Fig. 9). It is felt that further increase in x-ray output will result from improvement in betatron injection techniques rather than from changes in the synchrotron mode of operation.

Two types of electron gun have been used. One is a standard gun which has been used with betatrons for several years in this laboratory. It is quite similar to a gun described by D. W. Kerst.[14] A rather different gun, designed by J. P. Blewett, and initially studied on a rubber model, has been in use for a short time. The required orbit contraction is less than that in the earlier guns. Initial comparative experiments indicate that this improved gun gives a larger output than the standard gun.

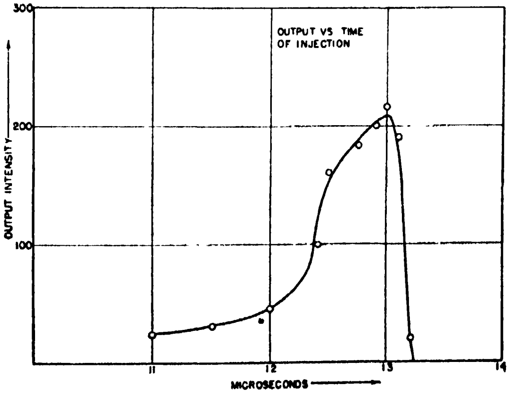


FIG. 8. Output *vs.* time electrons are injected by gun.

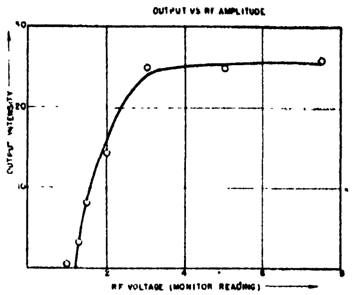


FIG. 9. Output *vs.* r-f accelerating volts. Monitor readings are directly proportional to peak voltage at the accelerating gap.

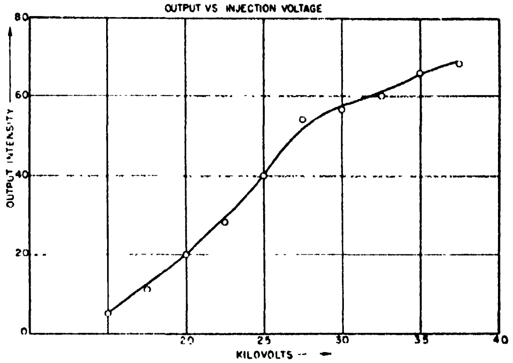


FIG. 10. Output *vs.* gun injection voltage.

The electron gun is always mounted on a side arm of the tube through a metal bellows which allows the gun position to be easily varied. Maximum x-ray output always seems to occur when the gun is located at a position in the tube just outside that where the magnetic field falls off inversely with the radius. It can be shown that in this region the electron beam has no radial stability, and accordingly the electrons can rapidly spiral in towards the center of the tube. A typical plot of x-ray output *vs.* injection voltage is shown in Fig. 10.

The x-ray output can be considerably increased by slight changes in the azimuthal distribution of the magnetic field. This is accomplished by placing flat sector-shaped coils along the top and bottom pole pieces of the magnet. These coils are connected to variable resistors. When the machine is operating, the output can usually

---

[14] D. W. Kerst, Rev. Sci. Inst. **13**, 387 (1942).

FIG. 11. Oscillogram of x-rays from internal target. The length of the sweep is 1000 microseconds. The trace begins at the time the r-f was turned off, when the electron orbit starts to contract. The irregular signal is from the output of a photo-multiplier tube placed in the emergent x-ray beam.

be increased several-fold by varying the current flowing through these coils.

When the x-ray intensity is low, the output of the machine is observed by a Geiger counter. With higher output, it is convenient to monitor with a 931 photo-multiplier tube contained in a light-tight box, and set a few feet from the target in the center of the beam, which is approximately 3° wide. The 931 is followed by a single stage of amplification and a cathode follower whose output is observed on an oscilloscope. The circuit has good frequency response which permits details of the x-ray beam structure to be observed on a fast sweep of the oscilloscope. An ionization chamber and d.c. amplifier are used to measure the average output of the machine.

The structure of the x-ray beam as a function of time is rather complicated. The distribution of x-rays from a contracted electron beam, as determined by a photo-multiplier tube is shown in Fig. 11. Upon moving the target by means of a metal bellows, it was found that a contracted electron beam gives a single x-ray pulse if the target is placed at a radius where $n>0[n=-d(\log B)/d(\log r)]$. If the target is at a smaller radius than that at which $n=0$, the x-ray beam becomes very complicated, probably because the electron beam has no vertical stability in regions where $n<0$.

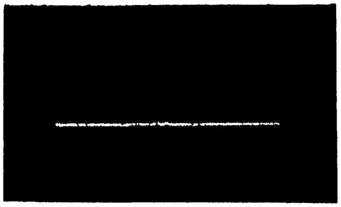


FIG. 12. Oscillogram of x-rays generated by electrons expanding to the gun as the magnetic field decreases. The 1000-microsecond trace is keyed at the time of r-f turn off. Two sharp pulses are registered by the photo-multiplier tube, about 50 microseconds apart.

An expanded electron beam, striking the gun as a target, gives a single pulse unless the gun is placed at a radius larger than that at which $n=1$. In this case, the x-rays emerge in two well separated sharp pulses as shown in Fig. 12. This splitting of the expanded beam is not understood at present, but is being investigated further.

In these experiments, the ejection of the electron beam to a target is accomplished by turning off the r-f resonator and letting the beam expand or contract as the magnetic field changes. If an x-ray pulse of shorter duration is needed for experimental work, some faster method of electron ejection should be used.[15]

## ACKNOWLEDGMENTS

The authors wish to acknowledge the many helpful suggestions made by Dr. E. E. Charlton, Dr. W. F. Westendorp, and Dr. J. P. Blewett. Mr. P. C. Noble, Mr. J. F. Tracy, Mr. R. L. Watters, and Mr. G. M. Kirkpatrick designed and tested many of the r-f components. Mr. F. Haber, Mr. G. N. Knowlton, and Mr. S. Sebis participated in much of the construction. This project has been supported throughout by the Office of Naval Research under Contract No. N5ori-178, task I.

---

[15] J. S. Clark, I. A. Getting, and J. E. Thomas, Phys. Rev. 70, 562 (1946).

# Cathode-Ray Presentation of Three-Dimensional Data*

OTTO H. SCHMITT
*Airborne Instruments Laboratory, Inc.,*** *Mineola, New York*

(Received March 24, 1947)

The cathode-ray oscilloscope is usually regarded as a means for displaying data in one or two variables as a function of time. By means of simple transformations which are easily performed electrically, it is possible to present three variable electrical data in the form of isometric or other conventional projections or as true perspective drawings. It is further possible to change the observer's viewpoint in the presentation coordinate system at will by turning range, elevation, and azimuth controls.

A more elaborate but similar set of transformations permits presentation of separate pictures to the two eyes. These pictures are optically superimposed but differ in such a way as to yield stereoscopically correct perspective pictures. These pictures are fully acceptable to the eye as patterns in space.

WHEN we are interested in examining an object of moderate size, we instinctively try to position it so that we look directly at it with both eyes from a distance ranging from a few inches to a few feet, depending upon the size of the object. We next turn the object or move about it so as to get several different views of it. Then, if we find certain details of the object particularly interesting, we move up and examine these details from a very short distance, sometimes even using a lens system to assist our eyes in focusing.

By the time we have completed such an examination we know pretty thoroughly what the object is like three-dimensionally, even though we have, in reality, seen only a series of pairs of the two-dimensional images formed on the retinas of our eyes.

The "realness" and "solidity" of an object in our field of view depends on a complex of subjective and objective factors, all of which our minds quickly and accurately put together into a mental image of the three-dimensional object. This mental picture is never a simple accumulation of the various retinal images: it is a real three-dimensional picture, nicely painted onto a framework provided by clues contained in the image pairs.

In this mental picture, missing details are neatly filled in, compromises are automatically made between various inconsistencies actually present in the data as seen, and errors resulting from optical and mechanical imperfections in the eye mechanisms are smoothed over.

Our visual apparatus is remarkably adaptable; if furnished with a few clues—even if some of them are contradictory—it will piece together a plausible, single clear picture, though this picture may not be the right one. When a wrong cue is taken and the data gets just too confusing, the eyes will try to find another picture that also fits the available data. The search for an intelligible picture will be repeated again and again before the eyes finally give up and register a blur.

This picture synthesis is the basis of many familiar optical illusions, especially those in which surfaces suddenly reverse or shift position. It is this process also that allows a skilled draftsman to convey a clear three-dimensional picture by a drawing on a flat sheet, using one of the systems of projection or perspective. In addition to having no real depth, these drawings may violently distort certain angles or lengths; yet so long as a few strong clues are given and the observer is trained to reject conflicting data, a clear picture can be formed.

In any scheme that proposes to present a three-dimensional picture to the eye in terms of one or a pair of flat pictures, we must attempt to suppress, or eliminate by education, the tendency for the eye to cling to features of the picture which reveal that it is "really" two-dimensional and we must emphasize those elements which

---

* Paper presented at the national convention of the Institute of Radio Engineers, March 3, 1947.

** Now at University of Minnesota, Department of Physics.

give the actually false impression of depth. Perception of depth by accommodation is subordinate to depth perception by binocular convergence, and depth perception by binocular convergence is subordinate to that by perspective. Consequently, a presentation that provides only perspective is convincing, whereas one that provides both perspective and stereoscopic view is very "real" indeed.

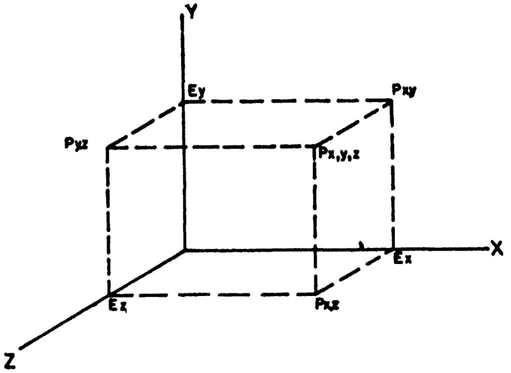


FIG. 1. Right-handed Cartesian coordinate system.

A cathode-ray tube provided with suitable deflecting and modulating voltages is capable of painting *any* two-dimensional picture within reasonable limits of brilliance, color, and grain size. It thus becomes obvious that a single cathode-ray tube can convey any three-dimensional-appearing picture that a perspective drawing can. In addition, two cathode-ray pictures, properly interrelated, can yield a real three-dimensional view if eye accommodation and comparison with surroundings are properly suppressed so that the eyes will be forced to rely on convergence and object shape for depth perception.

It is the purpose of this report to point out how easily the underlying principles of projective and perspective drawing and of stereoscopic photography can be applied to cathode-ray presentation so as to convert raw electrical data directly into vividly recognizable three-dimensional pictures.

Let us consider first the cases where both eyes look at the same picture. These are the projections and the perspective pictures familiar to every engineer and draftsman. In all these methods of presentation, a degree of three-dimensionality is given to the picture by having it resemble, in greater or less degree, the actual picture formed in the eye when viewing the object from some particular viewpoint, imaginary or real, the viewpoint in each case being chosen to bring out certain features of the picture.

In perspective drawing, some advantageous viewpoint is chosen and a serious attempt is made to develop, as accurately as is possible with simple rules of procedure, a picture like that which one eye would see if placed at that point. Projections, on the other hand, attempt to present as many as possible of the lines and angles of the picture in directly measurable form. Consequently, perspective drawings are easily recognized as three-dimensional even by the uninitiated, whereas projections tend to be distorted and often tax the eyes' ability to synthesize them into clear mental pictures, though great ingenuity has been shown in devising special projections that tend to neutralize the intrinsic distortions by counter-distortions.

To investigate the feasibility of presenting three-dimensional data on the cathode-ray tube let us start with the most general assumption: that we have at each instant three coordinate electrical data representing a point in our three-dimensional picture space, and that we also have a voltage specifying the spot brightness desired at that point, thus representing in a limited sense an additional parametric variable. We thus leave unrestricted the paths of the cathode-ray spot and are therefore able to deal with stationary or moving patterns made up of lines, spots, or surfaces; we need not consider separately whether the final picture is formed by scanning or by directed motion of the cathode-ray beam. (Remember, however, that direct, complete, three-dimensional scanning is ordinarily impractical because of the immense number of picture elements involved.)

Though coordinates other than Cartesian might obviously be used, we shall restrict ourselves to these for simplicity (Fig. 1). Let us call the three component voltages with which we are supplied $E_x$, $E_y$, and $E_z$, and let us think of $X$ as directed to the right, $Y$ as up, and $Z$ as toward us—the three axes comprising a right-handed Cartesian system. The three component voltages taken together represent a point $P$ in the three-dimensional $X$, $Y$, $Z$ space.

Plotting $E_x$ against $E_y$, $E_x$ against $E_z$, or $E_y$ against $E_z$ by means of a cathode-ray tube, we find that we get nothing new: we have one or another conventional cathode-ray presentation of part of the data. But looked at from the draftsman's standpoint, we have made standard multi-planar orthographic projection drawings: respectively, an elevation, a plan, and a profile projection of $P$ on these three projection planes. Following this lead we undertake to produce the draftsman's more difficult but more illustrative presentations on the cathode-ray tube.

Each of the orthographic projections is based on projection of the "object" data on a single projection plane by rays perpendicular to that plane. It can easily be shown that there is a perfectly general and always linear transformation from the three-dimensional voltage components $E_x$, $E_y$, and $E_z$ to the two-dimensional projection components on any specified projection plane. This transformation can be performed in the following manner.

Establish a new coordinate system $X'$, $Y'$, $Z'$ (Fig. 2) in such a fashion that one of its coordinate planes, say the $X'Y'$ plane, coincides with the desired projection plane. Call the cosines of the angles which the new $X'$ axis makes with the old $X$, $Y$, and $Z$ axes, respectively, $l_1$, $m_1$, $n_1$, and the cosines of the angles made by the new $Y'$ and $Z'$ axes with the old axes, correspondingly, $l_2$, $m_2$, $n_2$ and $l_3$, $m_3$, $n_3$. The general transformation for rotation of axes now yields the three component voltages $E_{x'}$, $E_{y'}$, $E_{z'}$, resolved along the new prime axes:

$$\begin{aligned}E_{x'}&=l_1E_x+m_1E_y+n_1E_z,\\ E_{y'}&=l_2E_x+m_2E_y+n_2E_z,\\ E_{z'}&=l_3E_x+m_3E_y+n_3E_z.\end{aligned}$$

Because each of these represents the total resultant voltage along one of the new prime axes, any pair represents the projection on one of the coordinate planes.

Electrically the meaning of this is simple. If we take, with a potentiometer, a fraction $l_1$ of our incoming $E_x$ voltage, add a fraction $m_1$ of $E_y$ and $n_1$ of $E_z$, apply this to our cathode-ray tube as $X$ deflection voltage, and use the corresponding fractions $l_2$, $m_2$, and $n_2$ of the original voltages for $Y$ deflection voltage, our tube will automatically plot the projection of the data on a new arbitrarily rotated $X'Y'$ plane. Since all the direction cosines are constant for any specified projection and since sine-cosine potentiometers, which are now readily available, yield the cosine directly when only the angle is set on the dial, this transformation is really a very simple one. It is often further simplified in particular resolutions by the fact that some of the angles are 0° or 90° and hence have cosines of 1 or 0—in which event no potentiometer is needed, all or none of a particular component voltage being used.

A specific example will serve to illustrate the simplicity of this projection transformation. Isometric projection occurs when the object axes are all inclined at equal angles of 35° 16′ to the projection plane. If the $Y$ projection is chosen to fall along the new $Y'$ axis, the transformation is:

$$\begin{aligned}E_{x'}&=\frac{\sqrt{2}}{2}(E_x-E_z)\\ &=0.707(E_x-E_z),\\ E_{y'}&=\sqrt{\frac{2}{3}}\left(E_y-\frac{E_x}{2}-\frac{E_z}{2}\right)\\ &=0.816\left(E_y-\frac{E_x}{2}-\frac{E_z}{2}\right).\end{aligned}$$

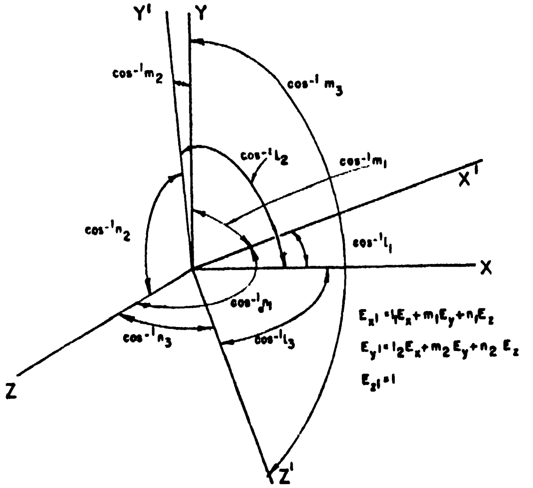


FIG. 2. General transformation for rotation of coordinates.

A little study will give values for the three coefficients for the $X'$ and for the $Y'$ that will yield any of the dimetric or trimetric projections.

Some of the most effective drawings are done in various "oblique projections." These are not really projections in the simple sense but they make use of the eyes' compromising ability to permit use of very simple drawing rules while still retaining good metric properties.

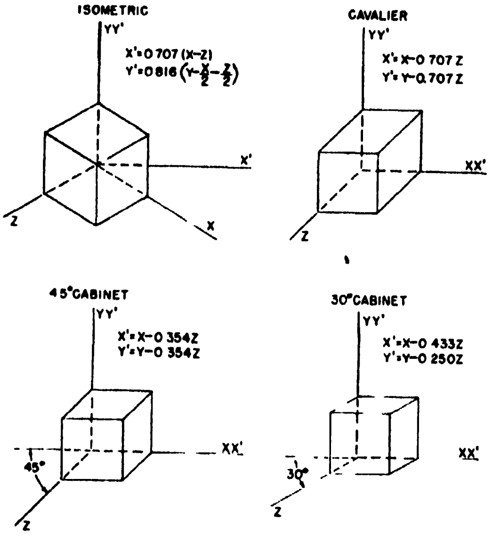


FIG. 3. Cube in isometric and various oblique projections with electrical transformation for each type of projection.

Figure 3 illustrates several projections, with a sketch of a cube in each, and gives the corresponding electrical transformations. Working independently, Carl Berkley of the DuMont Laboratories has devised and investigated certain of these projections in detail, both theoretically and experimentally, and has obtained excellent results. They were reported at the national convention of the Institute of Radio Engineers on March 5, 1947, and I shall therefore not elaborate on them here.

In military problems, in radar problems, indeed in most cases other than purely geometrical ones, we are likely to know angles other than those from which the direction cosines are derived. The most common case will be that of Fig. 4, in which voltage data are presented in rectangular $X$, $Y$, $Z$ coordinates. We wish (1) to take up our viewpoint at some little distance from the phenomenon, in a direction defined by an azimuth angle $\phi$ and an elevation angle $\theta$; (2) to see the data presented as a projection on a normal plane passing through a chosen origin; and (3), to keep our projection plane so oriented that the picture is still level, i.e., so that the new $X$ axis in the projection plane still lies in the old $XZ$ plane.

As it turns out, the transformation of electrical data to this new and completely general projection is very simple. Using two two-gang potentiometers, we can provide two knobs, one directly calibrated in azimuth and another marked in elevation, which will allow us to set the viewpoint anywhere around a complete sphere.

The transformation is simple if done as two separate coordinate rotations in proper order, as follows (Fig. 5). Starting as usual with data given as $E_x$, $E_y$, $E_z$, transform this for a rotation in azimuth $\phi$ about the $Y$ axis to give the $E$ data in terms of new coordinates $X'$, $Y'$, $Z'$:

$$E_{x'} = l_1E_x + m_1E_y + n_1E_z,$$
$$E_{y'} = l_2E_x + m_2E_y + n_2E_z,$$
$$E_{z'} = l_3E_x + m_3E_y + n_3E_z,$$

where $l_1$, $m_1$, $n_1$ are the direction cosines of the new $X'$ axis with respect to the old axes and $l_2$, $m_2$, $n_2$ and $l_3$, $m_3$, $n_3$ are the direction cosines of the new $Y'$ and $Z'$ axes. This yields:

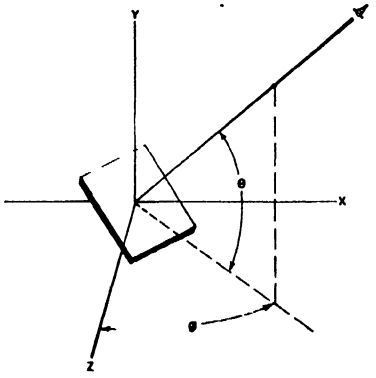


FIG. 4. Projection plane for viewpoint at azimuth $\phi$ and elevation $\theta$.

$$E_{x'} = E_x \cos\phi - E_z \sin\phi,$$
$$E_{y'} = E_y,$$
$$E_{z'} = E_x \sin\phi + E_z \cos\phi.$$

Now rotate this new coordinate system to a double-prime position (Fig. 6), where the $Z''$ axis, which is the desired line of view, is swung to an elevation angle $\theta$ by a rotation about $X'$.

$$E_{x''} = E_{x'}$$
$$E_{y''} = E_{y'} \cos\theta + E_{z'} \sin\theta$$
$$E_{z''} = -E_{y'} \sin\theta + E_{z'} \cos\theta.$$

To obtain these data, we need from the azimuth control system:

$E_x \sin\phi$ and $E_x \cos\phi$
(sine-cosine potentiometer No. 1),

$E_y$ (already available),

$E_z \sin\phi$ and $E_z \cos\phi$
(sine-cosine potentiometer No. 2).

For the second transformation, to put in elevation we need:

$E_{x'}$ (already available),
$E_{y'} \sin\theta$ and $E_{y'} \cos\theta$
(sine-cosine potentiometer No. 3),
$E_{z'} \sin\theta$ and $E_{z'} \cos\theta$
(sine-cosine potentiometer No. 4).

Since we desire to show the projection on the $X''Y''$ plane, we put the $E_{x''}$ and $E_{y''}$ voltages on the oscilloscope deflection plates and therefore do not need the $E_{z''}$ voltage. (Notice, however, that no extra potentiometers are required to obtain $E_{z''}$ because sine and cosine can come from a single unit. We shall have reason to return to $E_{z''}$ later, in connection with perspective and stereoscopic presentations.)

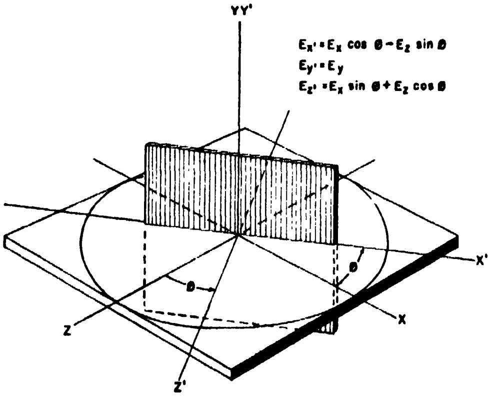


FIG. 5. Transformation by two separate coordinate rotations (rotation for azimuth angle $\phi$).

All the transformations thus far considered result in ideal or in stylized projections that are useful where distances along certain axes must be linearly measurable. Their extensive use in drafting as compared with perspective drawings results largely from their metric properties and from the difficulty of making good perspective drawings, especially from arbitrarily defined viewpoints. In the electric case, the use of view-rotating potentiometers permits any viewpoint to be chosen. For a complete presentation, therefore, it remains only to solve the problem of displaying perspective convergence on the cathode-ray tube.

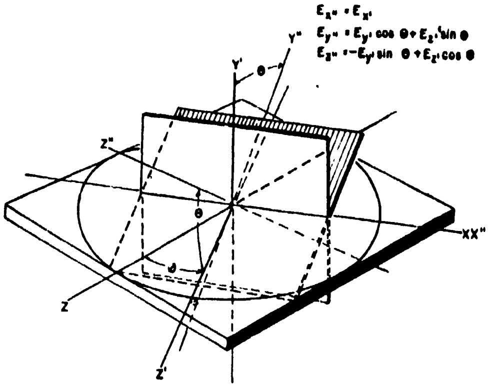


FIG. 6. Transformation by two separate coordinate rotations (rotation for elevation angle $\theta$).

Return to the original case—where $E_x$, $E_y$, and $E_z$ potentials are supplied directly or, what is equivalent, where the necessary rotations have already been accomplished—and consider how the picture would appear in perspective, looking along the $Z$ axis toward the origin (see Fig. 7). All points along the $Z$ axis project as a point at the origin $O$ on the $XY$ projection plane. Viewed from a distance $d$, the projection for a point $x = a$, $y = b$, $z = c$ would fall a distance $A = ad/(d-c)$ from the axis in the $X$ direction and $B = bd/(d-c)$ in the $Y$ direction. (This follows at once from inspection of similar triangles.) Since $a$ and $b$ are represented by the deflection voltages $E_x$ and $E_y$, respectively, and $d$ is a constant distance set on a dial, the multiplication indicated in the numerator can be accomplished by two simple linear potentiometers ganged to the "distance-of-view" control shaft. These potentiometers multiply the incoming $E_x$ and $E_y$ voltages, respectively, by the $d$ set on the dial. $c$ is a constantly varying quantity proportional to $E_z$. The reciprocal $1/(d-c)$ is obtained by using a non-linear amplifier. The $d$ in the denominator presents no difficulty because it is merely a d.c. bias proportional to the distance setting and is obtained from a third linear potentiometer on that shaft. Several components can be used in common for both $A$ and $B$ since both use identical circuits and contain the same multiplying factor $d/(d-c)$. This system has not been completed, and its performance characteristics cannot yet be stated.

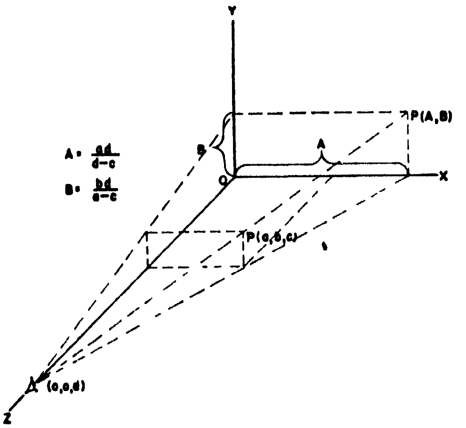


FIG. 7. Perspective projection presentation.

Adjustment of the distance-of-view control changes $d$—that is, it moves the observer's perspective viewpoint nearer or farther from the apparent object; since there is no focus problem here, it is equivalent to an ideal telephoto lens of adjustable magnification and corrected perspective.

Let us now consider what control over presentation we have, utilizing this perspective control combined with the azimuth and elevation controls previously described in connection with the orthogonal projection presentation. Let us take up a viewpoint a long way off and look over the data from all sides, as one does in inspecting an object visually. This we do by setting our perspective control for a long-distance viewpoint—just long enough, however, to permit us to see the whole picture at once. We now rotate the object—even though it may be actual radar data representing cubic miles of volume—by turning the two angular-selection controls. This permits us to select the one best viewpoint over the entire sphere of possible viewpoints. We now move ourselves nearer at will, going in among the data if this seems advantageous.

A word of caution should be injected at this point. It is feasible to utilize all the freedom just mentioned, but two factors must be remembered. On advancing the viewpoint into the field of data, data behind the observer may appear before him as a "negative." The occurrence of this phenomenon depends upon the design of the reciprocal amplifier and may in special cases require use of a special blanking circuit. However, this blanking circuit turns out to be quite simple.

The second factor concerns the apparent transparency of objects. It is obvious that no data can be presented which are not furnished electrically and, conversely, that all the data provided will be presented unless specifically blanked out. All "pictures" made by this process will then be "glass-block" pictures in which all surfaces of the object can be seen "through." This is generally desirable, especially when graphical or other line data are being presented. It may lead to optical illusions in certain cases, however, where the perspective picture shown is equivalent to another perspective picture of some other data configuration. Recall that in these pictures the eye is being asked to perceive depth solely by convergence of picture lines and cannot rely even on binocular convergence for a clue to real depth.

To investigate whether this one powerful aid to three-dimensional perception can be made to aid us in our attempt to present three-dimensional data realistically, let us see what further resolution of data would be needed to provide two suitably different sets of perspective data to be viewed by the two eyes separately to give the exact equivalent of stereoscopic, or perhaps even of enhanced stereoscopic, relief.

Ideal stereoscopic viewing presumes that both eyes look directly at the object from slightly separated points. In the language of our two

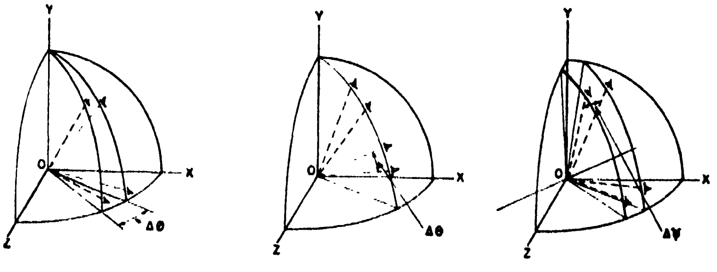


Fig. 8. Stereoscopic eyepoint separation by variation of $\phi$, $\theta$, and $\psi$.

angular-resolution potentiometers, this means that the tubes presenting data to the two eyes must be supplied with similar data for two slightly different viewpoints, and that this difference will depend upon the distance of the viewpoint from the data and upon the degree of stereoscopic enhancement desired.

As our coordinate transformation has been carried out, it is very simple to provide different viewpoints if those viewpoints differ by some simple and preferably constant amount in either $\phi$ or $\theta$. An extra set of contacts on the potentiometers that establish these angles or, at worst, a duplicate set of potentiometers lagging at a properly set angle will give the additional viewpoint.

Unfortunately, though the equivalent eye separation provided by a small change $\Delta\phi$ in azimuth (Fig. 8) corresponds to a desirable *horizontal* separation of eyepoints, it varies in amount with the cosine of the elevation angle $\theta$, since all the "longitude lines" converge.

Eyepoints produced by deviation of the elevation angle $\theta$ by a small amount $\Delta\theta$ are not subject to this error because lines of latitude are uniformly spaced. However, since they correspond to viewpoints above one another, the observer must view the picture as though lying on his side.

This problem can be ideally solved by a third re-resolution of the picture—this time by rotation about the $Y'''$ axis—to give eyepoints that are usable for any azimuth or elevation without readjustment. This is not a complicated transformation; it requires only two additional potentiometers and is therefore justified in special cases.

$$\begin{aligned}E_{x'''} &= E_{x''}\cos\psi - E_{z''}\sin\psi,\\ E_{y'''} &= E_{y''},\\ E_{z'''} &= E_{x''}\sin\psi + E_{z''}\cos\psi.\end{aligned}$$

These equations can also be expressed as follows:

$$\begin{aligned}E_{x'''} = {} & \cos\psi(E_x\cos\phi - E_z\sin\phi)\\ & -\sin\psi[-E_y\sin\theta + \cos\theta(E_x\sin\phi + E_z\cos\phi)],\end{aligned}$$

$$E_{y'''} = E_y\cos\theta + \sin\theta(E_x\sin\phi + E_z\cos\phi),$$

$$\begin{aligned}E_{z'''} = {} & \sin\psi(E_x\cos\phi - E_z\sin\phi)\\ & +\cos\psi[-E_y\sin\theta + \cos\theta(E_x\sin\phi + E_z\cos\phi)].\end{aligned}$$

If one does not need such complete generality but still wishes the advantages of stereoscopic vision, a good approximation can be had with very simple apparatus. Even the sine-cosine resolving potentiometers become unnecessary.

Consider the case where both eyes look at a nearby object $P$ on the axis of vision as projected

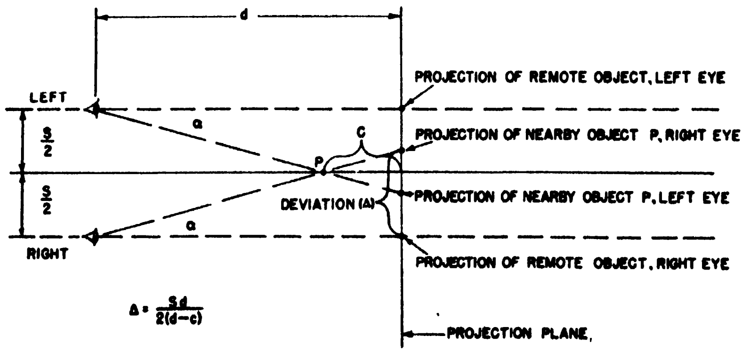

d
LEFT
S/2
S/2
a
a
C
P
DEVIATION (Δ)
RIGHT
PROJECTION OF REMOTE OBJECT, LEFT EYE
PROJECTION OF NEARBY OBJECT P, RIGHT EYE
PROJECTION OF NEARBY OBJECT P, LEFT EYE
PROJECTION OF REMOTE OBJECT, RIGHT EYE
PROJECTION PLANE.
Δ = Sd/2(d−c)

Fig. 9. Variation of binocular convergence angle with object distance.

on a plane perpendicular to the axis (Fig. 9). The convergence angle $\alpha$ for this point is determined solely by the proximity of the object to the eyes. The projection that represents this convergence must move to the right for the left eye and to the left for the right eye. The amount of deviation of each projection for an object to appear at a distance $d-c$ is given by

$$\Delta=\frac{Sd}{2(d-c)},$$

where $\Delta$ represents a deflecting voltage for application to the two cathode-ray tubes as $X$ bias of opposite polarities.

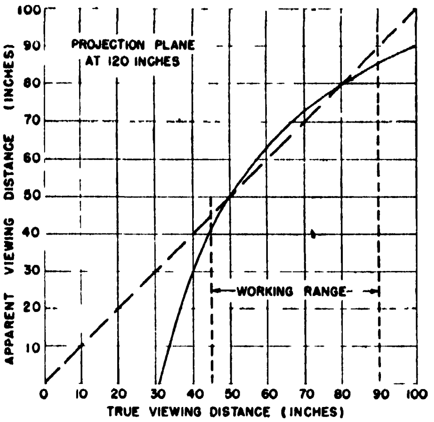


FIG. 10. Working range of simplified stereoscopic system.

$S$ is a d.c. voltage representing interocular distance and can be regarded as constant; $d$, as it multiplies $S$, can be represented exactly by a linear potentiometer on the distance-of-view shaft, working as a voltage divider on $S$. The factor of 2 is, of course, a mere scale factor.

Since $c$ is an instantaneous variable, it represents a reciprocal operation and requires a tube circuit; $d$, as it appears inside the parentheses with $c$ in the denominator, is applied as a d.c. bias to $c$. This quantity $(d-c)$ is already available from the perspective circuits, leaving only a reciprocal of a voltage to be taken. As before, this is accomplished by means of a non-linear tube circuit.

As a surprisingly good practical approximation in the case where the object field is not too deep or too close to the observer, a fraction of $(d-c)$, taken with reversed polarity and with a bias added, can be substituted for the reciprocal $1/(d-c)$. This apparently poor approximation works well because an error in the $\Delta$-voltage merely makes the object point appear stereoscopically a little nearer or farther than intended according to the perspective; this is an error that the eye can easily bear, as evidenced by its ability to understand perspective pictures on a flat sheet, where there is no correction at all. Figure 10 compares the apparent distance with the actual distance, using this simple stereoscopic system for a projection screen ten feet away and an object at various distances along the axis.

For objects off the axis the simplified stereoscopic method introduces small additional errors that increase as the object gets farther from the axis of view; however, these errors become important only for uncomfortably large angles of view.

If desired, the stereoscopic correction may be applied to simple projection data without the perspective correction. In this case the illusion of depth remains good, but objects seem to collapse as they approach and to expand as they recede in a very surprising manner.

In all these three-dimensional presentations the illusion is greatly enhanced if a few reference lines are sketched in to give the eye a comparison mesh with which to orient itself. There are many ways in which this can be done; one of the simplest is that of picture switching within the persistence period of the eye. It is relatively easy to commutate either mechanically or elec-

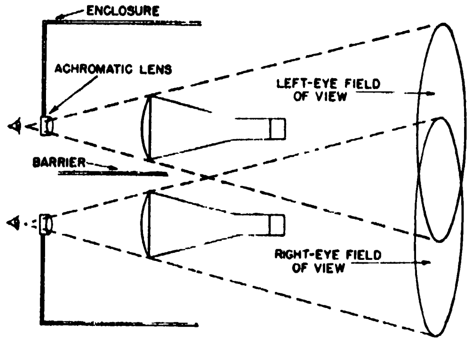


FIG. 11. Simple method for viewing stereoscopic images.

tronically between a sweep-circuit system and the picture data, and thus to draw in lines in any chosen direction and with any chosen spacing.

Though it is perfectly possible to superimpose such lines on the data, this may be difficult electronically in comparison with switching.

Again, it may be desired to measure the position of some intersection or other point in the picture space. This can be accomplished by a cursor system, using a calibrated set of controls to run a point or a line in a chosen orientation back and forth in the picture space.

The appearance on the market of multi-gun cathode-ray tubes offers an ideal solution to this problem of sketching in contour or grid lines, since it eliminates the need for any switching: with a two-gun tube, one gun paints in the data picture, and the other draws in the grid lines. Should one care to use the four-gun tube, the circuit is greatly simplified. Three of the guns can draw in three orthogonal coordinate-system grid lines in any of the conventional systems—Cartesian, polar, cylindrical—or even of the non-orthogonal systems.

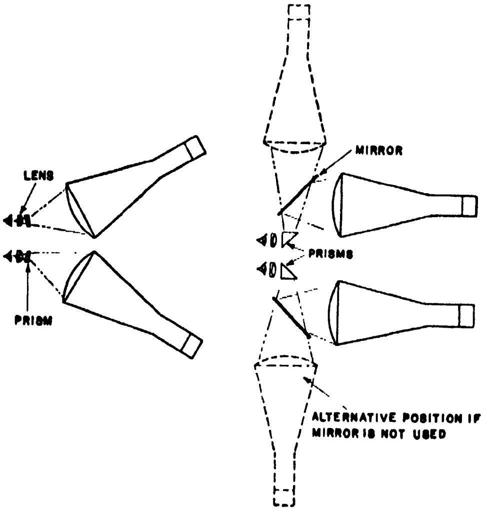


FIG. 12. Methods for viewing large cathode-ray tubes for stereoscopic presentation.

No special precautions need to be taken in viewing any of the monocular projections or the perspective pictures, whether they are viewed with one or with both eyes, since the eyes of a person used to reading such drawings are already completely conditioned to being deceived in these particular ways.

When viewing pictures that are mentally known to represent large objects, a little illusion is added by fitting the observer with positive lenses of sufficient strength to place the viewed image just short of infinity. Seeing the face of the cathode-ray tube and its mount does not weaken the illusion appreciably as it becomes part of the "drawing board."

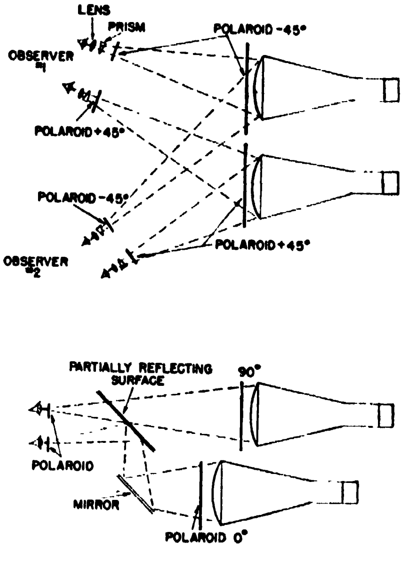


FIG. 13. (a) Use of Polaroid glasses to permit several observers to view stereoscopic presentation simultaneously; (b) Stereoscopic presentation for use in aircraft.

In contrast, data for stereoscopic viewing should be carefully presented to preserve the strong depth illusion of which this method is capable. The strength of stereoscopic reality found in the pictures varies markedly from observer to observer; airplane pilots, for example, are especially adept at perceiving it. That perception can be developed by learning is evident from the fact that a new observer seeing the stereoscopic CRO images for the first time may, if he is one of those least able to perceive the pictures, see almost no three-dimensional form to the images; however, after one hour of study, he will be able to see almost all the features. A very perceptive person will see all there is to see at first glance and will very likely be pointing out faults in the perspective rendition within five minutes.

Several satisfactory methods of viewing the stereoscopic images have been worked out. The simplest of these uses two small CRO tubes of diameter less than the interocular distance (Fig. 11). These tubes are viewed through simple achromatic lenses placed close to the eyes. The lenses make the tube face appear both at a distance and enlarged. An enclosure blocks out surrounding objects, and a partition prevents each eye from seeing the other's picture. This makeshift arrangement gives good results but permits seeing only such detail as can be presented on a three-inch tube. It also has the disadvantage that the eyes must be placed close to the viewing lenses.

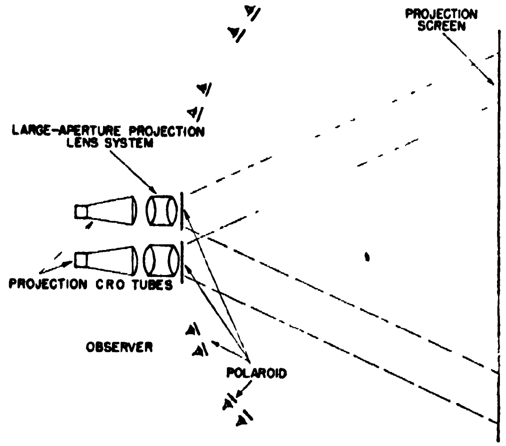


Fig. 14. Polarized projection for stereoscopic display.

If larger cathode-ray tubes are to be used, some optical arrangement must be made whereby each eye sees the center of one of the tubes directly ahead while the tubes are actually separated more than the interocular distance. Two of the easier ways of accomplishing this separation with prisms or mirrors are shown in Fig. 12.

By wearing Polaroid glasses it becomes possible to view the cathode-ray tubes over a reasonably wide angular range; consequently, several observers can see the presentation simultaneously (Fig. 13). If only the scopes are to be viewed, the observers can be fitted with prisms and enlarging lenses as well as Polaroids so that they can comfortably look at two tubes side by side and see them as superimposed because of the combined effect of the prism divergence and the Polaroid sorting out of the data. If this is undesirable, as in the case of data for an airplane pilot, the pictures can be shown superimposed with the aid of partially reflective surfaces. The Polaroid glasses are still worn but they do not interfere with ordinary vision.

Because of the selective reflection and transmission of plane polarized light by transparent sheets, proper choice of polarization plane will result in surprisingly little loss of light.

As a final and more elegant method, we may use polarized projection as a display means. Modern CRO projection techniques can provide a display several feet on an edge—a bright display that can be viewed by dozens of observers, each wearing ordinary 45° polarized Polaroid glasses (Fig. 14). By using highly directive projection screens, relatively little light is lost, and both images are displayed simultaneously on the same screen by simple polarization projection.

It is not my intention to enumerate here the many applications of three-dimensional cathode-ray presentation that can be made. I shall, however, list a few, chosen from widely separated fields, to illustrate its versatility.

The most general type of application is undoubtedly that in which three related quantities all vary simultaneously and where it is desired to visualize the whole complicated process. This visualization becomes especially important where the three interrelated quantities vary jointly as a function of still another parameter.

A case of this kind familiar to all of us is the common vacuum-tube characteristic. In its simplest three-variable form we have the co-variation of grid voltage, plate voltage, and plate current as one or more of them change in some chosen manner. Going to the more complicated case, we can present plate-current, plate-voltage, and grid-voltage data for a pentode or tetrode while we slowly change another parameter, say suppressor or screen voltage.

From the field of internal combustion or steam engineering we can take the case of the thermodynamic cycle. Here we can plot pressure, volume, and temperature together.

Acoustic engineers will, I am sure, find it instructive to plot the reactive and resistive components of loudspeaker impedance against frequency, thus giving in one easily photographed picture a good instantaneous summary of the device's over-all performance.

From the field of biophysics I would cite the advantages of plotting the entire set of heart potentials from the standard three cardiograph leads as a single space figure whose shape would characterize the patient's heart action. Similar thinking leads to a corresponding electro-encephalographic application.

For studying a.c. phenomena, both low frequency and high, the device is unusually useful. Non-linear components can be made to trace their current-voltage or reactance-resistance characteristics as a function of frequency, temperature, or other variable at will. Simultaneous plots of both complex impedance components as a function of frequency are a familiar need in the field of matching networks and filter design.

A three-dimensional cosmic-ray telescope could be built whose principal cost would be the set of Geiger counters.

The close relationship between the development of the theory of the three-dimensional oscilloscope and the theory of drafting leads one to wonder whether it might not profitably be put to work in the drafting room.

For presentation of radar- and sonar-type data, provision must obviously be made for supplying information in proper sequence, since it is impractical to scan all three-dimensional space. Thus, if we are scanning in elevation and azimuth by radar, we must be prepared, at the instant when the $x$ and $y$ coordinates correspond to the elevation and azimuth of a target, to supply the range to the $z$ circuit. Practicable pulse repetition frequencies do not permit scanning the $z$ axis in addition to the $x$ and $y$ axes. In the radar case, the advantages of the telescopic feature and the rotatable viewing position become obvious: for example, we could choose an optimum viewpoint to see whether planes were near a collision course.

Though it is true that three-dimensional television data can be transmitted by this system, real advantages would accrue only if the original data contained range information directly. Otherwise regular double-channel stereoscopic television has the advantage of simplicity.

From the antenna designer's viewpoint, one can see the elegance of the system as applied to presentation of a whole three-dimensional antenna radiation pattern visible in space, to be examined from any direction at will.

As a last example, I should like to propose the use of the system for exploration of electrostatic or magnetic fields. In this case a probe would be moved about in the field, and a "vector" on the oscilloscope screen would point out the direction and magnitude of the field at successive positions.

# A Least-Square Application to Relaxation Methods

O. L. Bowie
*Watertown Arsenal, Watertown, Massachusetts*

(Received April 24, 1947)

A least-square application to relaxation techniques is presented wherein the mean-square error is minimized in considering any corrections to a given set of values for the unknown function. A systematic procedure is, therefore, devised for weighting errors at various points of the lattice. The process has general applicability to various forms of equations and definitions of corresponding difference equations. As illustrations, Laplace's and the biharmonic equations are considered.

THE method of "relaxation" has recently come into prominence as a mathematical tool for overcoming the analytical difficulties encountered in the solution of partial differential equations from highly irregular shapes. Essentially, the differential equation is replaced* by a difference equation which may be solved either algebraically or by various iterative procedures. The iterative methods of solving the difference equation are known as relaxation techniques. It can be shown that if the mesh of the lattice used to cover the given region is made sufficiently fine, the solution of the difference equation approaches the solution of the original problem, and the difference quotients converge to the corresponding derivatives of the solution.

## I. A LEAST-SQUARES PROCEDURE

Consider the class of partial differential equations whose solutions have the additive property. That is, if $u_0$ and $u_1$ are solutions, then $(c_0u_0+c_1u_1)$ is also a solution where $c_0$ and $c_1$ are arbitrary constants. The same property will also hold for the solutions of the corresponding difference equations.

Let the given region be covered by a lattice. Then for defined points on the lattice, the differential equation can be replaced by a difference equation of the form

$$L(u_i)=0, \qquad (1)$$

where $u_i$ are the values of the function at the points of the lattice and $L(u_i)$ is a relationship valid for each $u_i$, which, in general, involves linear combinations of the $u_i$. If $L(u_i)=0$, for all $u_i$ of the lattice, and if the set of $u_i$ satisfies the original boundary conditions set up in a modified difference equation form, then the set of $u_i$ is a solution of the difference equation. In general, for a trial set, $\bar{u}_i$, it will be found that

$$L(\bar{u}_i)=p_i\neq 0. \qquad (2)$$

The $p_i$ are called the residuals and are defined for each point of the lattice. The relaxation procedure consists in altering the values of $u_i$ until $p_i\equiv 0$ for all $p_i$. In actual practice, sufficient accuracy is obtained by reducing the $p_i$ so that they are all nearly zero.

It should be noted that from the nature of $L(\bar{u}_i)$ an adjustment in $\bar{u}_i$ at a point will affect the values of $p_i$ at several points. The usual relaxation schemes (such as the Liebmann process) rely on smaller changes to the adjacent $p$'s as compared with the correction of $p_i$. It can be seen that the rate of convergence of relaxation methods will be speeded up if the correction of a $u_j$ is based on surrounding $p_i$ as well as on $p_j$. The following method which can be used for this purpose will now be presented.

Let the measure** of error of a given set of $\bar{u}_i$ over the whole lattice be defined as

$$\textstyle\sum_i [L(\bar{u}_i)]^2, \qquad (3)$$

where the summation extends over all the points of the lattice. The approximate solution, $\bar{u}_i$, will be chosen to satisfy the given boundary conditions. Now consider a second set of values, $\langle u_i\rangle$, such that $\bar{u}_i+c\langle u_i\rangle$ obeys the boundary conditions for any constant $c$. The question is now raised as to what value of $c$ will make the

---

* See reference 3.

** See reference 5.

total error a minimum for $\bar{u}_i+c\langle u_i\rangle$. Substituting into (3), one obtains for the total error

$$\sum_i [L(\bar{u}_i)+cL(\langle u_i\rangle)]^2. \qquad (4)$$

Differentiation with respect to $c$ yields for a minimum

$$c=-\sum_i L(\bar{u}_i)L(\langle u_i\rangle)/\sum_i [L(\langle u_i\rangle)]^2. \qquad (5)$$

For the criterion on the best correction in the sense of (3) to add to a single $\bar{u}_j$, one would choose a set of $\langle u_i\rangle$ identically zero at all points except at the single point corresponding to $\bar{u}_j$. The correction would be given by (5) if $\langle u_j\rangle=1$. Similarly, the best uniform correction to add to a group of the values of the function can be determined by (5). In this latter case, the set $\langle u_i\rangle$ again must be chosen so that $c\langle u_i\rangle+\bar{u}_i$ satisfies the boundary conditions for any arbitrary $c$. In addition, by defining $\langle u_i\rangle$ as identically equal to one at all points at which corrections are to be made and, otherwise, as identically equal to zero, the uniform correction, $c$, can again be determined by (5).

The procedure indicated in (4) and (5) is restrictive in the sense that it provides for only a uniform correction to the given set of $\bar{u}_i$. This is evidently an unnecessary restriction, for one can choose several sets of $\langle u_i\rangle$ with several corresponding arbitrary constants and again derive conditions on the arbitrary constants by referring back to (3). This, however, would involve the solution of a set of simultaneous equations and the simplicity of the form of (5) would be lost.

## II. APPLICATION TO LAPLACE'S EQUATION

The Dirichlet problem is that of finding a solution of Laplace's equation

$$(\partial^2 u/\partial x^2)+(\partial^2 u/\partial y^2)=0 \text{ in } R, \qquad (6)$$

where $u$ has specified values on the boundary of $R$. After $R$ is covered with a square lattice, then in rectangular coordinates and in terms of neighboring points the difference equation corresponding to (6) becomes a set of relationships in the form

$$L(u_0)=u_1+u_2+u_3+u_4-4u_0=0, \qquad (7)$$

where the notation is indicated in Fig. 1. One manner in which the $\bar{u}_0$ are adjusted so that the forms (7) are satisfied is to add iteratively to the quantities $\bar{u}_0$ the correction $\frac{1}{4}L(\bar{u}_0)$. The corrected $\bar{u}_0$ will then satisfy (7), to be sure, but $L(\bar{u}_1)$, $L(\bar{u}_2)$, $L(\bar{u}_3)$, and $L(\bar{u}_4)$ will be increased by an amount $\frac{1}{4}L(\bar{u}_0)$.

To illustrate the principle outlined in the first section, the criterion for the best correction will now be derived in terms of (7). The function $\langle u_i\rangle$ will thus be defined as unity at a single point and identically zero elsewhere. By (7) it is seen that $L(\langle u_i\rangle)$ has the values $-4$ at the point corresponding to the functional value of unity, 1 at four neighboring points, and zero elsewhere. A glance at Eq. (5) will show that for a single correction to $\bar{u}_0$ the best in the sense of (3) will be

$$c=-\frac{1}{20}[L(\bar{u}_1)+L(\bar{u}_2)+L(\bar{u}_3)+L(\bar{u}_4)-4L(\bar{u}_0)]. \qquad (8)$$

Formulae similar to (8) can likewise be obtained for a single correction to several of the functional values.

For corrections of points adjacent to the boundary, it is necessary to define in some manner values for $L(\bar{u}_i)$ at boundary points in order to use (8). Such a definition is somewhat arbitrary as it imposes more conditions on the set of simultaneous difference equations than there are unknowns. At the same time a wise choice of the definition is important to insure rapid convergence of the iterative process. In most cases, it is sufficient to modify the form of

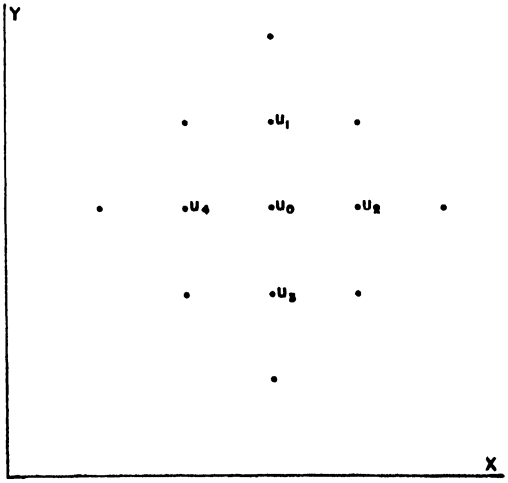


FIG. 1.

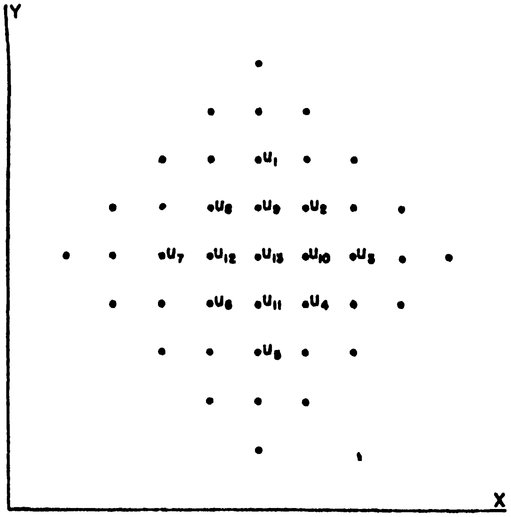


FIG. 2.

(7) by using forward differences when on the boundary. Then, during the process of iteration, a stage will be reached when the computer can vary the fictitious boundary $L(\bar{u}_i)$'s by visual inspection to speed the convergence.

It is interesting to note that a single correction based on (8) weights the values of $\bar{u}_i$ at the thirteen points shown in Fig. 1. Since the least-squares correction is based on adjacent residuals as well as the residual at the given point under consideration, it has a tendency to "smooth" the pattern of residuals. In general, this property has an advantage over usual "point by point" procedures. However, in some situations, the averaging effect slows down the rate of convergence. When the residuals are either predominantly positive or negative, the convergence is greatly speeded up by the larger corrections in the values of the function which one makes by using the usual "point by point" procedures. In practice, therefore, to obtain the full advantage of the least-square procedure, corrections should be made to the function at several points. It was found that a uniform correction, derived by the least-squares criterion, for small blocks of points proved very simple to handle and insured rapid convergence of the solution. For use of the least-squares method in the solution of the Dirichlet problem, therefore, it is recommended that the uniform correction to simple-shaped blocks of points be used in conjunction with the single-point correction.

## III. APPLICATION TO THE BIHARMONIC EQUATION

The biharmonic problem involves the solution of the equation

$$[\partial^4 u(x,y)/\partial x^4]+2[\partial^4 u(x,y)/\partial x^2\partial y^2]+[\partial^4 u(x,y)/\partial y^4]=0 \text{ in } R, \quad (9)$$

subject to specified values of $u$ and certain of its derivatives on the boundary of $R$. In terms of a square lattice of points, the difference equations involve thirteen points for each point of the lattice. With the notation of Fig. 2, the difference equations are in the form

$$L(u_{13})=(u_1+u_3+u_5+u_7)+2(u_2+u_4+u_6+u_8)-8(u_9+u_{10}+u_{11}+u_{12})+20u_{13}=0. \quad (10)$$

From Eq. (5) it is seen that for a single correction to $\bar{u}_{13}$ the best in the sense of (3) will be

$$c=-\frac{1}{676}\{[L(\bar{u}_1)+L(\bar{u}_3)+L(\bar{u}_5)+L(\bar{u}_7)]+2[L(\bar{u}_2)+L(\bar{u}_4)+L(\bar{u}_6)+L(\bar{u}_8)]-8[L(\bar{u}_9)+L(\bar{u}_{10})+L(\bar{u}_{11})+L(\bar{u}_{12})]+20L(\bar{u}_{13})\}. \quad (11)$$

Similarly, formulae for corrections to blocks of values of the function can be derived. Definitions must be set up for values of $L(u_i)$ where the lattice points are boundary points or points adjacent to the boundary. These may be set up by use of forward differences as in the previous section and adjusted as the computation progresses.

To obtain the full advantage of the least-squares procedure in the solution of the biharmonic equation, it is recommended that the uniform correction to simple-shaped blocks of points be used in conjunction with the single-point correction.

## IV. DISCUSSION

The principle outlined in Section I can be applied to relaxation procedures involved in the solution of general types of equations of physics. The forms $L(u_i)=0$, themselves, may be defined in many ways. The difference equation may be set up using higher ordered differences, or differences using diagonal neighbors. A non-homogeneous equation can be likewise treated by suitably defining $L(u_i)=0$. In general, the form of $L(u_i)$ should be chosen from the standpoint of the amount of weighting of errors possible with a minimum of calculation.

Although only a single-point correction has been explicitly derived in Sections II and III, the process for deriving similar formulae for corrections to blocks of points has been outlined. It would seem, in general, that such formulae for blocks of points should be derived only for the case of a uniform correction. More elaborate considerations would necessarily complicate the form of the correction formula.

Preliminary tests of this least-squares procedure have been found quite satisfactory when combining uniform corrections to blocks of points with single-point corrections. The advantage over usual "point by point" procedures increases as the process continues.

## BIBLIOGRAPHY

1. G. H. Shortly and R. Weller, "The numerical solution of Laplace's equation," J. App. Phys. **9**, 334–348 (1938).
2. D. G. Christopherson and R. V. Southwell, "Relaxation methods applied to engineering problems," Proc. Roy. Soc. **A168**, 317–350 (1938). See also H. W. Emmons, "The numerical solutions of heat conduction problems," Trans. A.S.M.E. **65**, 607–615 (1943).
3. David Moskovitz, "The numerical solution of Laplace's and Poisson's equations," Quart. App. Math. **2**, 148–163 (July 1944).
4. I. S. Sokolnikoff, *Mathematical Theory of Elasticity* (McGraw-Hill Book Company, Inc., New York, 1946), Chapter V.

---

# An X-Ray Diffraction Study of the Silver-Magnesium Alloy System

HAROLD R. LETNER* AND S. S. SIDHU**

*Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania*

(Received April 29, 1947)

Four homogeneous solid solutions were found in the silver-magnesium alloy system. The $\alpha$-primary solid solution of magnesium in silver has a f.c.c. structure. The $\beta$-solid solution has a b.c.c. structure. The $\gamma$-solid solution appears to have a complicated structure. While others have identified it as hexagonal, no definite structure could be assigned to it from our diffraction data. The $\delta$-solid solution has a h.c.p. structure of magnesium.

The lattic parameter $a_0$ in the $\alpha$-solid solution increases with increasing magnesium concentration but exhibits a marked negative deviation from Vegard's Law. The $a_0$ in the $\beta$-solid solution, however, increases linearly with increasing magnesium concentration.

Silver and magnesium atoms are distributed at random in the $\alpha$-phase but take up preferred positions in the $\beta$-phase. The ordering process takes place throughout the $\beta$-phase, but is most nearly complete in the region of 50 atomic percent Mg.

## INTRODUCTION

THE Ag-Mg alloy system presents an interesting crystal structure problem in that Ag is known to have a face-centered cubic (f.c.c.) structure, while Mg is hexagonal close-packed (h.c.p.). Ag, atomic number 47, is monovalent and has an atomic radius of 1.44 angstrom units (A); Mg, atomic number 12, is divalent and has an atomic radius of 1.60A.[1] The purpose of the present work was to study by x-ray diffraction methods the changes which take place in the structure of Ag as Mg atoms are added to its lattice.

The phase equilibrium diagram of the Ag-Mg system was first determined by Zemczuznyj.[2] Subsequent investigations were carried out by Saeftel and Sachs,[3,4] Payne and Haughton,[5]

---

* Now at Mellon Institute of Industrial Research, Pittsburgh, Pennsylvania.

** On leave of absence now at Argonne National Laboratory, Chicago, Ill.

[1] W. Hume-Rothery, *The Structure of Metals and Alloys* (Chemical Publishing Company, Inc., Brooklyn, New York, 1939), Part III, pp. 30–49.

[2] S. F. Zemczuznyj, "Über die Leigerungen des Magnesiums mit Silber," Zeits. anorg. Chemie **49**, 400 (1906).

[3] F. Saeftel and G. Sachs, "Festigkeitseigenschaften und Struktur einiger begrenzter Mischkristallreihen," Zeits. Metallkunde **17**, 258 (1925).

[4] W. Guertler, "Neue Silberlegierungen," Zeits. Metallkunde **19**, 68 (1927).

[5] R. J. M. Payne and J. L. Haughton, "Alloys of magnesium, part IV.—The constitution of the magnesium-rich alloys of magnesium and silver," J. Inst. Metals **60**, 351–364 (1937).

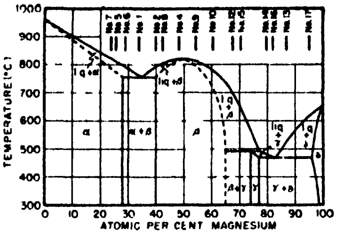
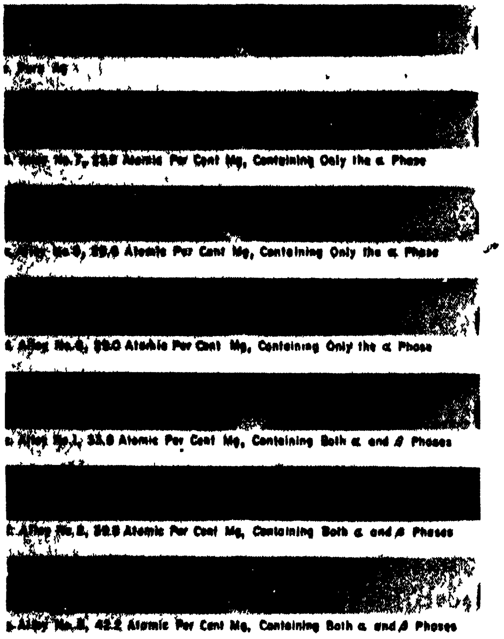


FIG. 1 Silver-magnesium phase equilibrium diagram showing compositions of the alloys studied.

Hume-Rothery and Butchers,[6] Owen and Preston,[7] Ageew and Kuznezow,[8] and George.[9] Figure 1 is the phase equilibrium diagram of Zemczuznyj as given by Hansen,[10] modified to include the more recent results of the other investigators for the Mg-rich portion of the system. The vertical lines with numbers 1 to 17 refer to this study.


a. Pure Ag

b. Alloy No. 7, [illegible] Atomic Per Cent Mg, Containing Only the α Phase

c. Alloy No. [illegible] Atomic Per Cent Mg, Containing Only the α Phase

d. Alloy No. 6, 28.0 Atomic Per Cent Mg, Containing Only the α Phase

e. Alloy No. 1, 33.8 Atomic Per Cent Mg, Containing Both α and β Phases

f. Alloy No. 2, 39.8 Atomic Per Cent Mg, Containing Both α and β Phases

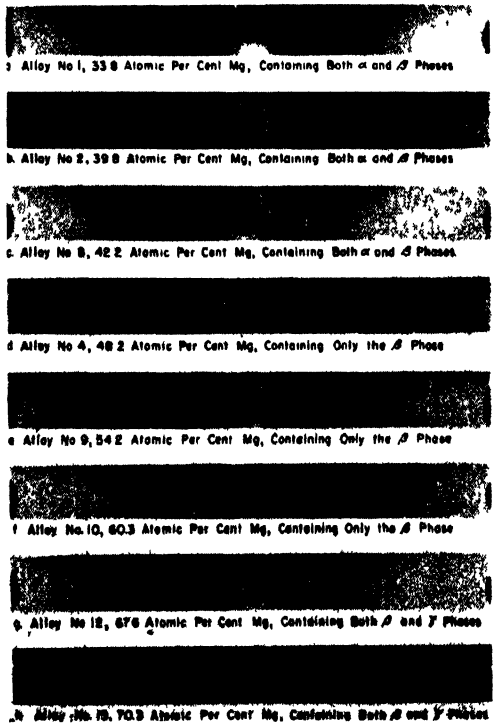
g. Alloy No. 8, 42.2 Atomic Per Cent Mg, Containing Both α and β Phases

FIG. 2. X-ray photograms of pure silver and alloys containing the α-solid solution of magnesium in silver. Camera diameter 57.3 mm, copper $K_\alpha$ radiation

## EXPERIMENTAL PROCEDURE

Fifteen alloys having compositions indicated in Fig. 1 were prepared in an automatically temperature-controlled electric furnace under a commercial flux obtained from the Dow Chemical Company. Upon completion of the melting process, the resulting ingots, weighing approximately 50 grams each, were allowed to cool in the furnace. The cooling rate in the solidifying range of the alloys was approximately 3C° per minute.

Samples of filings were removed from the centers of the ingots with specially cleaned iron files and placed in evacuated hard glass tubes. After annealing for five hours at 525°C to remove mechanical strains set up in the crystal lattice by the filing process, the samples were analyzed chemically for both Ag and Mg. Portions of the


a. Alloy No 1, 33.8 Atomic Per Cent Mg, Containing Both α and β Phases

b. Alloy No 2, 39.8 Atomic Per Cent Mg, Containing Both α and β Phases

c. Alloy No 8, 42.2 Atomic Per Cent Mg, Containing Both α and β Phases

d. Alloy No 4, 48.2 Atomic Per Cent Mg, Containing Only the β Phase

e. Alloy No 9, 54.2 Atomic Per Cent Mg, Containing Only the β Phase

f. Alloy No. 10, 60.3 Atomic Per Cent Mg, Containing Only the β Phase

g. Alloy No 12, 67.6 Atomic Per Cent Mg, Containing Both β and γ Phases

h. Alloy No. 13, 70.3 Atomic Per Cent Mg, Containing Both β and γ Phases

FIG. 3. X-ray photograms of alloys containing the β-solid solution of magnesium in silver. Camera diameter, 57.3 mm; copper $K_\alpha$ radiation.

---

[6] W. Hume-Rothery and E. Butchers, "The solubility of silver and gold in solid magnesium," J. Inst. Metals **60**, 345–350.

[7] E. A. Owen and G. D. Preston, "The atomic structure of silver-magnesium and gold-zinc," Phil. Mag. **2**, 1266 (1926).

[8] N. Ageew and V. G. Kuznezow, "Roentgenographic study of alloys of magnesium and silver," Bull. Acad. Sci. U.R.S.S., Classe sci. math. nat., Ser. chem., 289–309 (1937).

[9] H. H. George, *X-Ray Investigation of Silver-Magnesium Alloys* (Thesis) (University of Pittsburgh, 1939).

[10] M. Hansen, *Der Aufbau der Zweistoff-Legierungen* (Julius Springer Verlag, Berlin, 1936), p. 38.

same samples were used in making powder x-ray-diffraction photograms to insure that the compositions of the filings used for diffraction studies and those used for chemical analyses were identical.[11]

Powder x-ray diffraction photograms were made with filtered Cu $K_\alpha$ radiation using a camera of 57.3-mm diameter. Lattice parameters for the f.c.c. and b.c.c. structures were calculated from the relation:

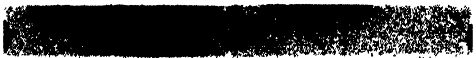
$$a_0=\frac{n\lambda}{2\sin\theta}(h^2+k^2+l^2)^{\frac{1}{2}},\qquad(1)$$

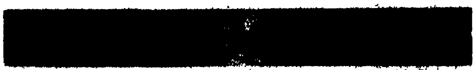
where $a_0$ is the lattice parameter (i.e., the cube edge of the unit cell), $n$ is the order of the diffracted ray, $\lambda$ is the wave-length of the radiation used, $h\,k\,l$ are the Miller indices of the planes contributing to the diffracted ray, and $\theta$ is the angle between the incident x-ray beam and the diffracting planes. Lattice parameters for the h.c.p. structure were calculated from the corresponding expression:

$$a_0=\frac{n\lambda}{2\sin\theta}(4/3(h^2+hk+k^2)+(l/c)^2)^{\frac{1}{2}},\qquad(2)$$

where $a_0$ and $c_0$ are the lattice parameters, and $c=c_0/a_0$ is the axial ratio.

Values of $\theta$ were obtained directly from the photograms by measuring the spans between corresponding lies of the diffraction pattern.[12] From these values the lattice parameters were calculated by means of Eqs. (1) or (2) and plotted against $\cot\theta$. The resulting curves were extrapolated to $\cot\theta=0$ to obtain final values of the lattice parameters, since the percentage error approaches zero as $\theta$ approaches 90°.[13-15]


a Alloy No. 12, 67.6 Atomic Per Cent Mg, Containing Both β and γ Phases


b Alloy No. 15, 70.3 Atomic Per Cent Mg, Containing Both β and γ Phases

c. Alloy No. 14, 79.8 Atomic Per Cent Mg, Containing Both γ and δ Phases

d. Alloy No 16, 82.0 Atomic Per Cent Mg, Containing Both γ and δ Phases

e Alloy No 13, 87.0 Atomic Per Cent Mg, Containing Both γ and δ Phases

f. Alloy No. 17, 95 1 Atomic Per Cent Mg, Containing Both γ and δ Phases

g Pure Mg

FIG. 4. X-ray photograms of alloys containing the γ- and δ-solid solutions of magnesium in silver and of pure magnesium. Camera diameter, 57.3 mm; copper $K_\alpha$ radiation.

The relative intensities of the diffraction lines were estimated and tabulated simultaneously with the lattice parameter measurements. Photograms of alloys containing the α- and β-phases were scrutinized for superstructure lines. For the α-phase, this amounted to a search for lines corresponding to reflecting planes having Miller indices $h\,k\,l$ which are mixed odd and even integers and for the β-phase, lines corresponding to planes for which the sums of the Miller

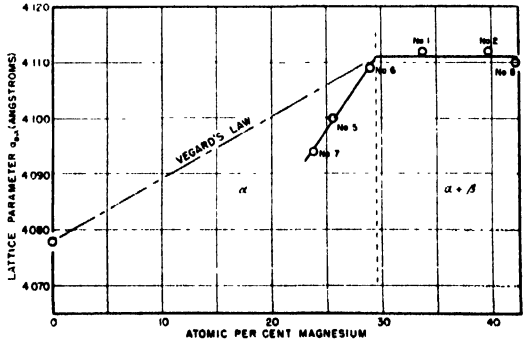


FIG. 5. Variation of lattice parameter $a_{0\alpha}$ with composition in α-phase silver-magnesium alloys.

---

[11] W. Hume-Rothery and P. W. Reynolds, "Application of the Debye-Scherrer method of x-ray crystal analysis," J. Inst. Metals **60**, 303–317 (1937).

[12] M. J. Buerger, "Design of x-ray powder cameras," J. App. Phys. **16**, 504 (1945).

[13] M. J. Buerger, "Precision determination of linear and angular constants of single crystals," Zeits. f. Krist. **97**, 433–468 (1937).

[14] Charles S. Barrett, *Structure of Metals, Metallurgy and Metallurgical Engineering Series* (McGraw-Hill Book Company, Inc., New York and London, 1943), pp. 134–138.

[15] B. E. Warren, "The absorption displacement in x-ray diffraction by cylindrical samples," J. App. Phys. **16**, 614 (1945).

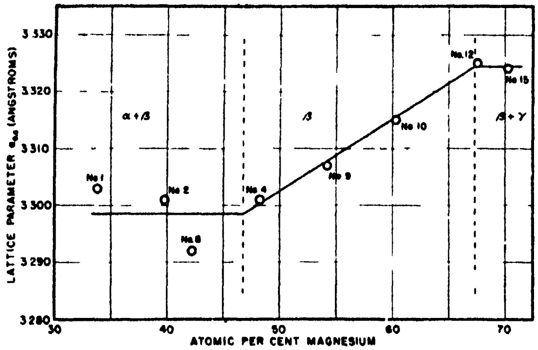


FIG. 6. Variation of lattice parameter $a_{0\beta}$ with composition in β-phase silver-magnesium alloys.

indices $h+k+l$ are odd integers.[16–18] No satisfactory analysis could be made for superstructure in the γ- and δ-phases because of the large number of extremely weak lines obtained on the photograms.

## EXPERIMENTAL RESULTS

### A. Phases and Phase Boundaries

Four types of crystal structure were found to exist in the Ag-Mg alloys investigated. These were a f.c.c. structure corresponding to the α-phase, a b.c.c. structure corresponding to the β-phase, an undetermined structure corresponding to the γ-phase and a h.c.p. phase belonging to the δ-phase. (See Fig. 1.) Attempts to determine the structure of the γ-phase were not successful; photograms of alloys containing this phase showed a multiplicity of very faint and indistinct lines. Ageew and Kuznezow have identified the structure as hexagonal.[8] X-ray photograms of the diffraction patterns corresponding to the various structures found are shown in Figs. 2–4.

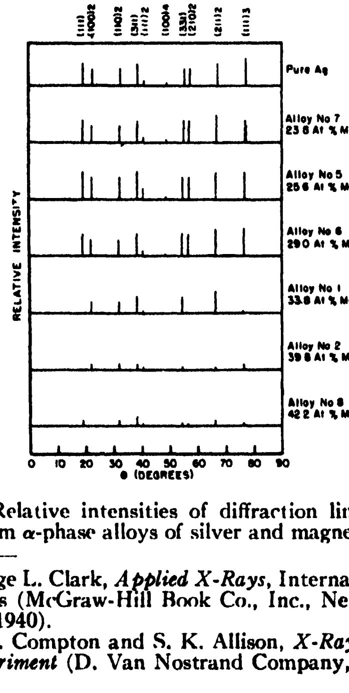


FIG. 7. Relative intensities of diffraction lines obtained from α-phase alloys of silver and magnesium.

Although no attempt was made to determine accurately phase boundaries, one point in this connection seems worthy of mention. Alloy No. 6 having a composition of 29.0 atomic percent Mg should, according to the phase equilibrium diagram in Fig. 1, contain both α- and β-phases. The photogram of this alloy, Fig. 2, shows only the f.c.c. lines of the α-phase. Similarly, alloy No. 8, also Fig. 2, having a composition of 42.2 atomic percent Mg, should contain only the β-phase according to Fig. 1. The photogram of this alloy shows unmistakably the f.c.c. and b.c.c. lines corresponding to both the α- and β-phases. Thus, while there is general agreement with the results of Zemczuznyj, the limited data available from the present work indicate that the

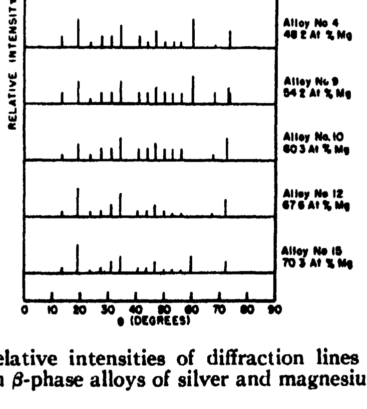


FIG. 8. Relative intensities of diffraction lines obtained from β-phase alloys of silver and magnesium.

---

[16] George L. Clark, *Applied X-Rays*, International Series in Physics (McGraw-Hill Book Co., Inc., New York and London, 1940).

[17] A. H. Compton and S. K. Allison, *X-Rays in Theory and Experiment* (D. Van Nostrand Company, New York, 1935).

[18] Charles S. Barrett, *Structure of Metals, Mettallurgy, and Metallurgical Engineering Series* (McGraw-Hill Book Company, Inc., New York and London, 1943), p. 127.

boundaries of the two-phase region between the $\alpha$- and $\beta$-phases are not straight lines of constant composition as shown in Fig. 1 but are slightly displaced toward the Mg-rich end of the diagram.

### B. Lattice Parameters

The lattice parameter $a_{0\alpha}$ varies with composition as shown in Fig. 5. The distribution of the alloy compositions studied does not permit the drawing of a complete curve for the entire phase; however, it is obvious that only a curve exhibiting a negative deviation from the linear relationship required by Vegard's law will fit the points obtained. This result is at variance with the work of Ageew and Kuznezow[8] who found that the lattice parameter increases linearly with increasing Mg content.

Within the precision of measurement, alloys of the $\beta$-phase show a linear variation of the lattice parameter $a_{0\beta}$ with change in composition. Rather wide discrepancies appear among alloys 1, 2, and 8 containing both the $\alpha$- and $\beta$-solid solutions, all of which should have the same value of $a_{0\beta}$. In Fig. 6 the value chosen for $a_{0\beta}$ in the two-phase region is the arithmetic mean of the values for the three alloys. The resulting phase-boundary discontinuity, indicated by the dashed line, lies well to the Mg-rich side of alloy No. 8 as previously mentioned.

The indistinctness of the diffraction lines on photograms containing the $\gamma$-phase rendered impractical any attempt to observe lattice parameter variations in alloys 13, 14, 16, and 17. The lattice parameters of pure Mg were found to be $a_0=3.202$A and $c_0=5.199$A, in agreement with accepted values.

### C. Superstructure

Photograms of alloys containing the $\alpha$-phase show no evidence of superstructure. Figure 7 shows the relative intensities of all the diffraction lines observed on the photograms of each of the $\alpha$-phase alloys. It will be noted that all of the lines correspond to reflections from planes which have either all odd or all even Miller indices; i.e., the (111), (100)2, (110)2, (311), (111)2, (100)4, (331), (210)2, (211)2, and (111)3 planes.

Photograms of the $\beta$-phase alloys, on the other hand, contain unmistakable evidence of superstructure lines. In Fig. 8 it will be noted that not only are lines corresponding to even values of the sum $h+k+l$ present, but relatively strong lines corresponding to odd values are also present. If the distribution of Ag and Mg atoms in the b.c.c. lattice were random, lines corresponding to the (100), (111), (210), (221), (311), (320), (322), and (410) planes would not appear on the photograms. Hence it is quite evident that the Ag and Mg atoms take up preferred positions in the $\beta$-solid solution to form an ordered structure. Comparison of superstructure lines corresponding to the various $\beta$-phase alloys reveals that they are strongest near the middle of the $\beta$-phase, indicating that the segregation of the Ag and Mg atoms into cube centers and cube corners is most nearly complete when equal numbers of each atom are available; i.e., in the region of 50 atomic percent Mg.

---

## Erratum: The Torsion of a Rubber Cylinder

[J. App. Phys. 18, 444 (1947)]

R. S. Rivlin
*National Bureau of Standards, Washington, D. C.*

Equations (4) should read

$$\Theta=2\psi r\left(\frac{\partial W}{\partial I_1}+\frac{\partial W}{\partial I_2}\right)$$

and

$$Z=-2\psi^2a^2\left(\frac{\partial W}{\partial I_2}\right)_{r=a}+2\psi^2\int_a^r r\left\{\left(\frac{\partial W}{\partial I_1}-2\frac{\partial W}{\partial I_2}\right)-2\left[\frac{\partial^2 W}{\partial I_1{}^2}+(3+\psi^2r^2)\left(\frac{\partial^2 W}{\partial I_1\partial I_2}\right)\right.\right.$$
$$\left.\left.+(2+\psi^2r^2)\frac{\partial^2 W}{\partial I_2{}^2}\right]\right\}dr+2\left[\frac{\partial W}{\partial I_1}-\left(\frac{\partial W}{\partial I_1}\right)_{r=a}\right]+4\left[\frac{\partial W}{\partial I_2}-\left(\frac{\partial W}{\partial I_2}\right)_{r=a}\right].$$

# Preliminary Analysis of Microwave Transmission Data Obtained on the San Diego Coast Under Conditions of a Surface Duct

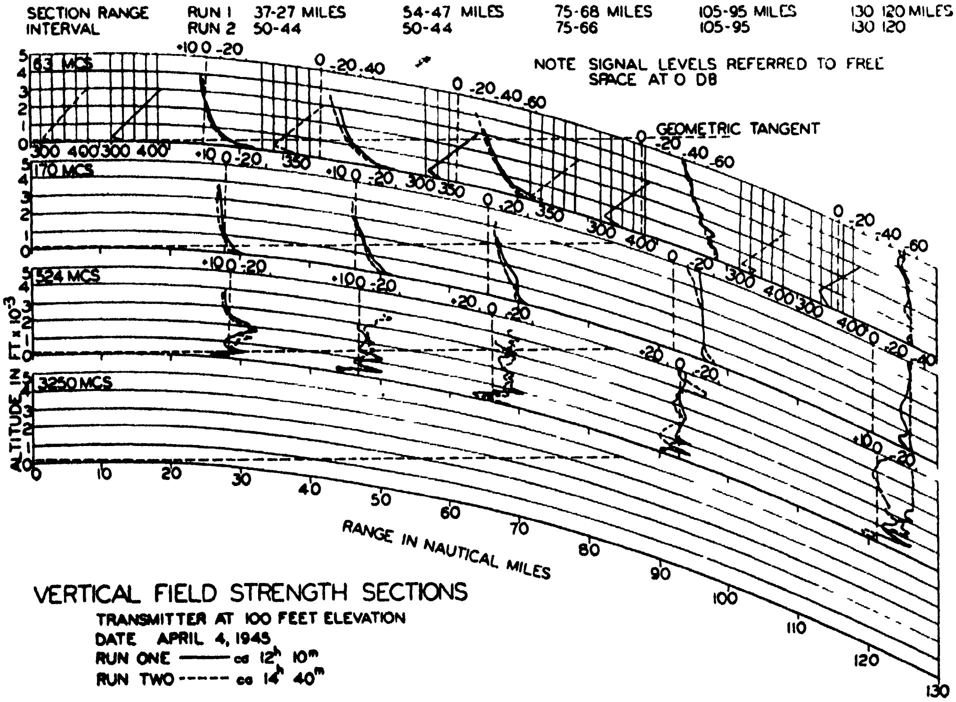
C. L. Pekeris* and M. E. Davis**

(Received May 6, 1947)

The transmission data for 63 Mc and 170 Mc obtained by the U.S.N.E.L. on the San Diego coast on April 4, 1945, when a strong duct was formed near the ground, were analyzed by wave theory. The average observed $M$-curve was fitted with a linear-exponential term, and the electromagnetic field beyond the horizon was computed by the theory of normal modes. Good agreement between theory and experiment was obtained for both frequencies at a range of 32 miles (see Figs. 7 and 8). In the ranges between 32 and about 70 miles the observed variation of intensity with height agrees with theory, but the theoretical horizontal decrement is less than the observed by about 0.1 d.b./n. mile for 63 Mc, and by about 0.2 d.b./n. mile for 170 Mc. Beyond about 80 miles the observed field tends to assume a uniform distribution with height and shows little horizontal attenuation. Such an effect had been observed by the U.S.N.R.L. expedition to Antigua for 9-cm waves under conditions of an oceanic surface duct. Its interpretation is still wanting.

## 1. INTRODUCTION

In this paper is given a preliminary analysis of transmission data for 63 Mc and 170 Mc obtained by the U.S.N.E.L. along the San Diego coast under atmospheric refraction conditions which produce a so-called surface duct. The set-



Note: Run 2 should read 37–27 instead of 50–44.

Fig. 1.

---

* Columbia University Mathematical Physics Group, on leave of absence from Massachusetts Institute of Technology. This paper is based on work done for the Naval Research Laboratory under contract N6ori-110, task order #3 with the Navy's Office of Research and Inventions. Publication assisted by the Ernest Kempton Adams Fund for Physical Research of Columbia University.

** U. S. Navy Electronics Laboratory, San Diego, California.

up of the experiment and the observed $M$-curves, as well as the measured cross sections of field intensity at various ranges from the transmitter are shown in Fig. 1. The average observed $M$-values based on 12 soundings are shown by curve $A$ in Fig. 2. Of the 12 soundings, the two taken at the maximum range of 115 miles deviate markedly from the others; another of the soundings taken about 2 hours before the experiment at a range of 10 miles also appears anomalous. This may be caused by both spacial and temporal variation of the $M$-curves. Leaving out these three soundings one obtains an average $M$-curve shown by curve $B$ of Fig. 2.

Taking the $M$-curve $B$ as characteristic of the average atmospheric refraction condition over the relevant portion of the transmission path, we attempt to represent it by a linear-exponential term. Various such representations are shown in Fig. 3, with curve $E$ appearing to give the best fit. This theoretical curve has a surface value for $(M-M_0)$ of 72.3, which is probably considerably in excess of the actual value. However, as has been pointed out by Professor Hartree, transmission is not seriously affected by the detailed distribution of $M$ near the surface because of the vanishing of the field intensity at the surface.

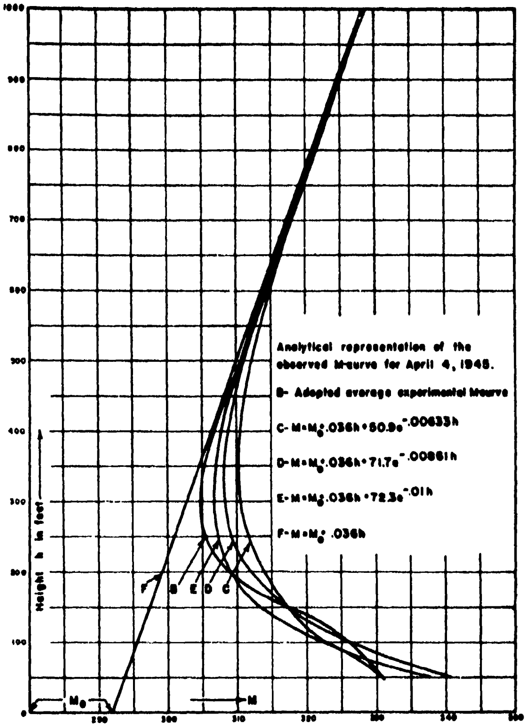


FIG. 3.

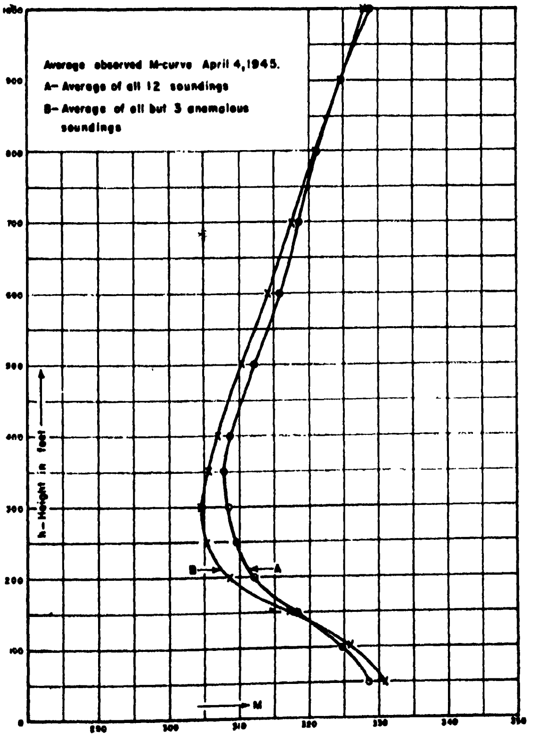


FIG. 2.

The surface duct under discussion, as represented by curve $E$

$$M = M_0 + 0.036h + 72.3e^{-0.01h}, \qquad (1)$$

with its surface $M$, excess of 72.3, and duct height of 300 feet is to be contrasted with curve $C'$ of the Antigua oceanic surface duct previously analyzed,[1]

$$M = M_0 + 0.036h + 12.06e^{-0.0714h}, \qquad (2)$$

where the surface $M$, excess, and duct height had values of only 12.06 and 45 feet, respectively. It is therefore of interest to test how wave theory can predict the electromagnetic field under such vastly different conditions. It turns out that when expressed in natural units, curve $E$ for 170 Mc (176 cm) is close to curve $C'$ for 9 cm, that is, the San Diego duct had about the same trapping power for 176 cm waves as the Antigua duct[2] had for 9 cm waves.

---

[1] C. L. Pekeris, Proc. I.R.E. **35**, 435 (1947).

[2] M. Katzin, R. W. Bauchman, and W. Binnian, Proc. I.R.E. **35**, 891 (1947).

TABLE I. Data for 63 Mc, $\lambda=476$ cm, $H=442$ feet, $L=25.8$ n. miles. Curves $C$, $D$, and $E$ are shown in Fig. 3.

| Curve | $\alpha$ | $\lambda$ | $D_1$ | Theor. decrement d.b./n. mile | Obs. decrement d.b./n. mile | Obs. decrement range miles |
|---|---|---|---|---|---|---|
| $C$ | 3.21 | 2.80 | $-1.03+1.82i$ | 0.61 | 0.79 | 32 71 |
| $D$ | 4.52 | 3.80 | $-1.02+1.88i$ | 0.63 | 0.73 | 32–100 |
| $E$ | 4.56 | 4.42 | $-0.99+1.98i$ | 0.67 | 0.66 | 71–100 |

## 2. ANALYSIS OF THE FIELD BY THE THEORY OF NORMAL MODES

According to the theory of normal modes, the field $FS$ relative to free space, due to a dipole radiating in a horizontally stratified atmosphere is given by[1]

$$FS=20\log_{10}\left[2(\pi x)^{\frac{1}{2}}\left|\sum_{m=1}^{\infty}e^{-A_m x+iB_m x}U_m(z_1)U_m(z_2)\right|\right]\text{d.b.}\quad(3)$$

Here $x$ denotes horizontal distance from transmitter expressed in natural units $L$, $z_1$, and $z_2$ the heights of transmitter and receiver above ground expressed in natural units $H$, where

$$L=3.304\lambda^{\frac{1}{3}}\text{ n. miles},\quad H=7.243\lambda^{\frac{2}{3}}\text{ feet},\quad(4)$$

and $\lambda$ is the wave-length in centimeters. The height-gain functions $U_m(z)$ and the characteristic constants $A_m$ and $B_m$ can be computed by known methods[2] for the case of the linear-exponential $M$-curve given in Eq. (1). These depend on the values of the parameters $a$ and $c$ in the expression

$$M=M_0+0.036h+ae^{-ch}.\quad(5)$$

Again, it is convenient to use instead of $a$ and $c$ the natural parameters

$$\alpha=3.85a\lambda^{-\frac{1}{3}},\quad \lambda^{***}=Hc,\quad(6)$$

whereby the exponential term in (5) is transformed into $\alpha e^{-\lambda z}$.

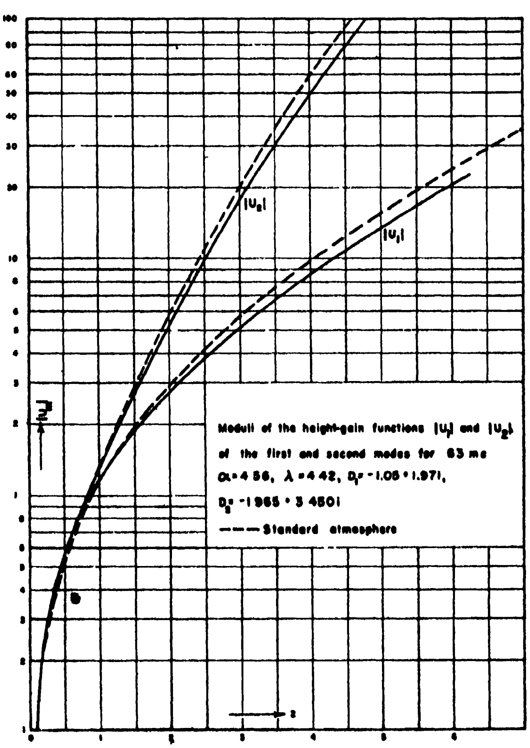


FIG. 4.

Tables I and II show the results of the analysis for the 63 Mc and 170 Mc, respectively. The characteristic values $D_1=B_1+iA_1$ of the first mode were obtained by interpolation from the values given in reference (3).[3] In the case of curve $E$ and 63 Mc the value of $D_1$ shown in Table I required extrapolation beyond the maximum value of 4 for $\lambda$ treated in reference (3). As a check, $D_1$ was recomputed by the perturbation method and a value of $-1.05+1.97i$ was obtained. This method also yielded a convergent value for $U_1(0)$ of $-0.043+1.17i$. With these constants the differential equation for the height-gain function of the first mode $U_1(z)$

$$\frac{d^2U(z)}{dz^2}+Q(z)U(z)=0,\quad Q(z)=z+D+\alpha e^{-\lambda z}\quad(7)$$

was integrated using an interval $\Delta z=0.1$ and the

TABLE II. Data for 170 Mc, $\lambda=176$ cm, $H=228$ feet, $L=18.5$ n. miles.

| Curve | $\alpha$ | $\lambda$ | $D_1$ | Theor. decrement d.b./n. mile | Obs. decrement d.b./n. mile | Obs. decrement range miles |
|---|---|---|---|---|---|---|
| $C$ | 6.23 | 1 44 | $-1\ 84+0.69i$ | 0.32 | 0.61 | 32–71 |
| $D$ | 8.77 | 1.96 | $-1.48+0\ 72i$ | 0.34 | 0.41 | 32 100 |
| $E$ | 8.84 | 2.28 | $-1.22+0\ 94i$ | 0.44 | 0.17 | 71–100 |

---

*** Not to be confused with the $\lambda$ in (4) which denotes wave-length.

[3] See a forthcoming paper by C. L. Pekeris and W. S. Ament. To be published in Phil. Mag.

recursion formula

$$\left[1+\frac{(\Delta z)^2Q_1}{12}\right]U_1=2U_0-U_{-1}$$
$$-\frac{(\Delta z)^2}{12}(10U_0Q_0+U_{-1}Q_{-1}),\quad (8)$$

which is correct to within $[(\Delta z)^2/240]\delta^4(QU)$, where $\delta^4$ denotes the fourth difference. The perturbation method also gives for the second mode $D_2=-1.965+3.450i$, $U'_2(0)=0.041+1.084i$, and these were used to obtain $U_2(z)$ by (7) and (8). The results for the first two modes are shown in Figs. 4 and 5. Clearly, for 63 Mc the trapping of the modes is small, so that for the application in view it would have been sufficient to use the solutions for a standard atmosphere.

In the case of 170 Mc and curve *E*, values of $D_1=-1.22+0.94i$ and $U'_1(0)=1.14+1.04i$ were obtained by interpolation from the "method II" values tabulated in reference (3). With these, the height-gain function $U_1(z)$ was integrated using (7) and (8), and the result is shown in Fig. 6. Here the trapping is considerable, the theoretical decrement being less than half the standard value. From the perturbation method one obtains the approximate values $D_2=-2.01+3.12i$, $U'_2(0)=-0.36+1.05i$.

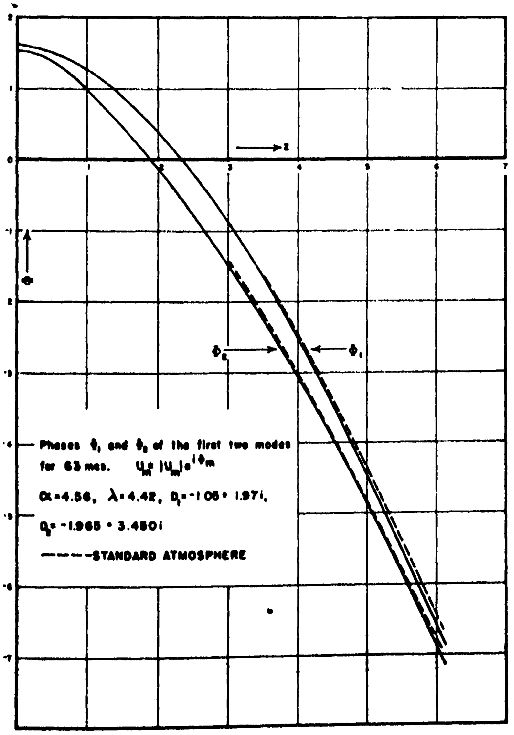


FIG. 5.

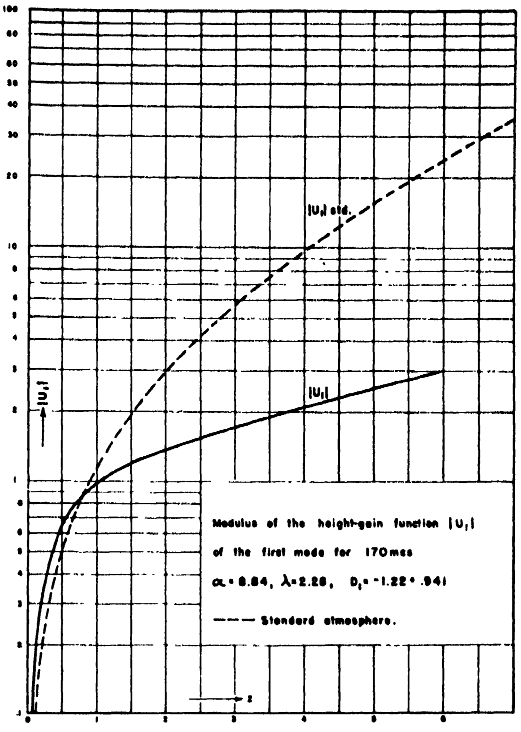


FIG. 6.

### 3. DISCUSSION OF RESULTS

Figure 7 shows a comparison of observed field intensities relative to free space with theoretical. The first two modes were used in the calculations, and where the second mode contributed more than 50 percent to the total field the theoretical curves were terminated. The observed increase of intensity with height, as well as the detailed distribution of intensity with height, shows satisfactory agreement with theory out to a range of 71 miles, but the observed decrement is greater than the theoretical (which is close to the standard value) by about 0.1 d.b./n. mile. In the case of 170 Mc shown in Fig. 8 there is again good agreement at a range of 32 n. miles, and satisfactory agreement in the variation of intensity with height out to 49 n. miles. The observed

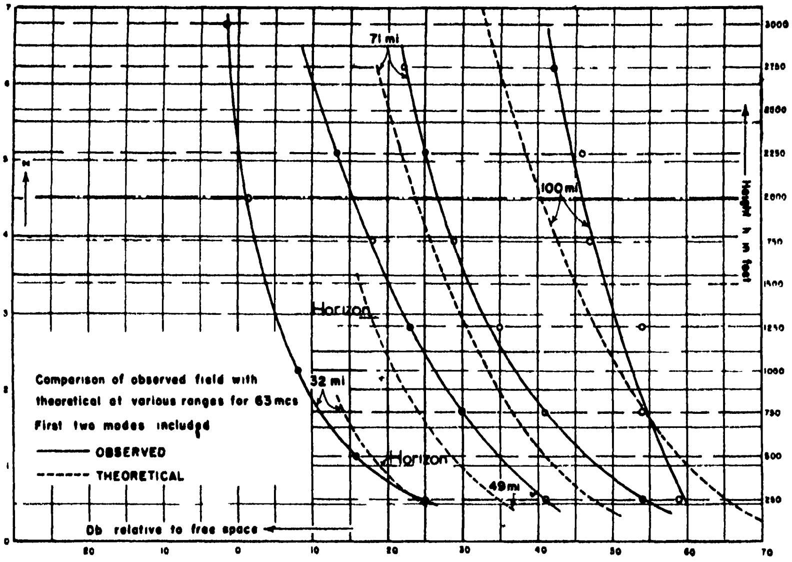


FIG. 7.

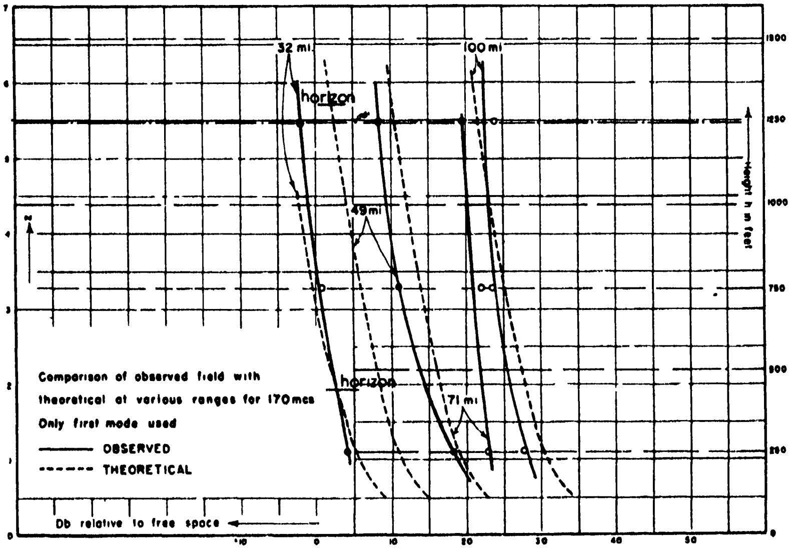


FIG. 8.

decrement, however, while only 0.7 of standard, exceeds the theoretical value by about 0.3 d.b./n. mile at this range.

It is possible that a weaker surface duct than the one adopted in this preliminary analysis would bring the theoretical decrements to closer agreement with observation, while not materially affecting the shape of the height-gain curve. Part of the excessive observed attenuation no doubt originates from scattering by inhomogeneities in the atmosphere. However, the tendency of the observed field to assume a uniform distribution with height, with little horizontal attenuation beyond about 70 miles, is probably an effect which cannot be accounted for by the present formulation of the normal mode theory.

# Letters to the Editor

## Modified Cavity Oscillator for the Generation of Microwaves

GEORGE G. BRUCK
*Specialties, Inc., Syosset, New York*
March 13, 1947

THE cavity consists of two co-axial cylinders and two end-plates. $N$ longitudinal slots in the wall of the inner cylinder are used to create the proper type of field. This layout resembles the well-known "turbator." A qualitative picture of the field can be given easily although an exact calculation is rather laborious. Figure 1 shows a probable aspect related to the $\pi$-mode of the magnetron. Figure 2 is a probable configuration similar to the $TE_{01}$ mode of a cylindrical guide.

Electrons may be made to describe circular orbits in the region between the two cylinders by applying a positive potential to the inner one. As in the orbital beam tube (RCA) there is no need for a magnetic field. Because of progressive sorting, $N$-electron clouds are formed. This is explained in the following way: the fringing field near the slots has an electric component in the direction of the flight-path of the electrons. This component of the field accelerates or decelerates the electrons. The accelerated ones, which absorb energy from the field, fly outward, while the decelerated ones, which give up energy to the field, approach the inner cylinder because of loss in velocity. At the following slot the unfavorable electrons, the faster ones, pass through a weaker fringing field than the favorable ones. They extract less energy from the field than the favorable ones are adding to it. The effect obtained is quite pronounced because the intensity of the fringing field declines exponentially with the distance from the slot. The in-phase electrons spiral inward. The out-of-phase electrons spiral outward. There exists an added tendency of proper phasing because of the difference in path length for the fast and slow electrons. The angular displacement required to give proper phasing can be progressively approached by the faster electrons because of the increasing radius of their orbit.

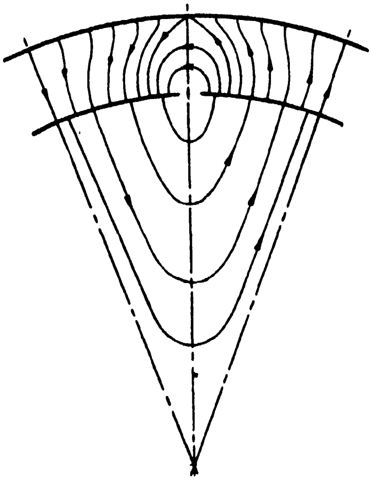


FIG. 1. Probable electric field $\pi$-mode.

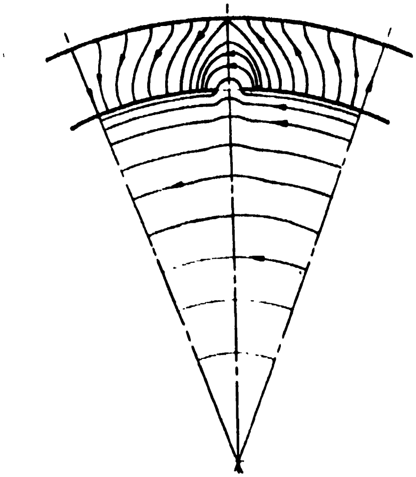


FIG. 2. Probable electric field corresponding to $TE_{01}$ mode.

The cathode and accelerating grid are both parallel to the cylinder axis. Their location is such as to create minimum interference with the orbits of the electron. There are no slots provided in their immediate neighborhood (Fig. 3). The voltage on the accelerating grid is related to the potential between the inner and outer cylinders in order to produce initially circular orbits.

Microwave power is extracted from the inner cavity by any approved method probe, loop, or window.

For modes of the $TE_{01}$ type, the cavity may be split along a plane, perpendicular to the axis. This permits the insertion of a dielectric vacuum-tight window and location of the electron stream in the sealed and evacuated part. The output iris and simple means for frequency adjustment are located in the non-evacuated section.

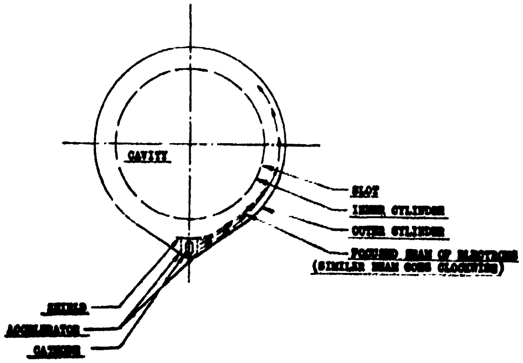


FIG. 3. Potential distribution: cathode, shield, outer cylinder . . . negative; accelerator, inner cylinder . . . positive.

No evaluation of the specific properties of this tube has as yet been made. It is anticipated that high output at higher frequencies will be obtainable, particularly above 30 kmc. It is further believed that this type of oscillator may be made to work at frequencies exceeding 100 kmc.

## Particle "Growth" in the Electron Microscope

V. E. COSSLETT*
*Cavendish Laboratory, Cambridge, England*
June 9, 1947

IT has recently been observed by Watson[1] that carbon black particles appear to increase in size when used as objects in an electron microscope. He rightly stresses the importance of this disturbing effect in particle-size determinations, and calls for an exchange of experience in order to clear it up.

We observed a similar effect some months since with the smallest zinc-oxide crystals obtainable by burning zinc in air in a coal-gas flame (see Fig. 1); they were being used as test specimens in investigating the performance of a Siemens electron microscope. We have now obtained similar results with zinc, molybdenum, and magnesium oxides in the RCA Type-B microscope. Figure 2 shows a typical series of photographs from zinc oxide supported on an RCA nickel grid. Plates (a), (c), and (e) were taken at intervals of 2 minutes after bringing the needles into the field of view, in critical illumination, i.e., with the source imaged on the specimen, which was thus exposed to maximum energy flux. Plate (b) is of a field lying within 10 microns of that first examined, and plate (d) is of a field lying on the opposite side of the grid hole containing these two fields, i.e., about 75 microns from them and outside the focused image of the source. It is clear that the effect is confined to the region covered by the latter, and its very immediate neighborhood.

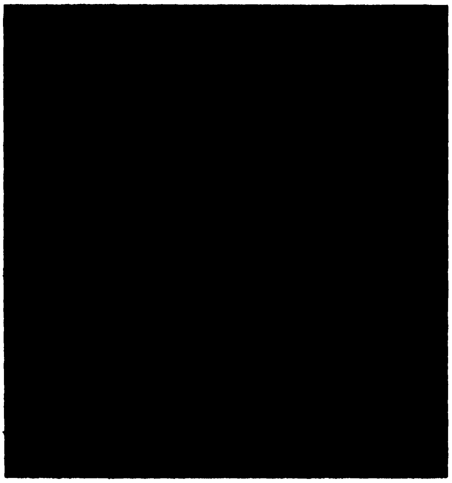


FIG. 1. Zinc-oxide smoke particles (20,000×) after *ca.* 3-minutes exposure to electron beam in Siemens electron microscope (on nickel grid).

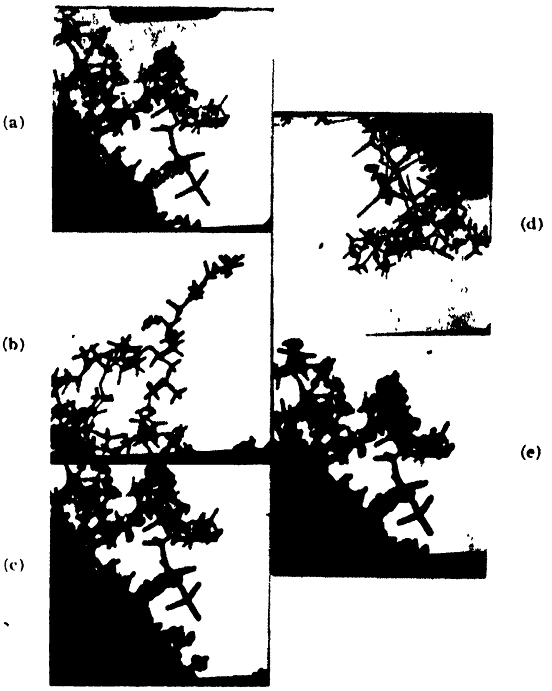


FIG. 2. Zinc-oxide smoke particles in RCA Type B electron microscope 12,000× (on nickel grid).

a. After *ca.* 2 minutes in beam.
b. Adjacent field to (a) (within 10μ), after *ca.* 3 minutes in beam.
c. Field (a) after a further 2 minutes in beam.
d. Field *ca.* 75μ distant from field (a).
e. Field (a) after a further 2 minutes in beam.

Observation of crystals supported on long chains of their like from the opposite side of a grid hole, or by adventitious, loosely-adhering material on the near side, shows that such growth does not occur when there is a poorly conducting path between specimen and grid. Watson found little effect on particles supported in the center of a grid space on a collodion film, but considerable effect after shadowing with chromium. We find little effect at all in such supported specimens, wherever they be with respect to the grid, and no enhancing influence of a thick gold deposit (*ca.* 100 atomic layers). This is of importance for the estimation of the size of virus and 'phage particles, which cannot be observed otherwise than on a supporting film. We have not been able to find any increase in size of gold-shadowed turnip virus particles, nor of the flagella of bacteria, after exposures in the beam up to 10 minutes.

In contradistinction to Watson, who suggests that the growth arises from the deposition of organic material from the vapor state in the form of ion clusters, we incline to the view that the effect is due to ejection of adsorbed matter, and perhaps of metal itself, from the grid wires under electron bombardment. The local temperature, since the beam image is of the order of 10 microns only in diameter, will be very high and may well reach 1000°C; in the Siemens microscope it is possible to fuse the nickel grid if the filament is a little over-run. Prior cleaning of the grids with organic solvents, and heating in the electron beam in vacuum before depositing the specimen, bring about a small improvement only, making it very doubtful that organic impurities on the grid are responsible. The fact that magnesium oxide, a highly refractory material, shows more rapid growth than either zinc or molybdenum oxides, suggests that it cannot be due to recrystallization or the growth of some crystals at the expense of others; in any case a diminution in size is never observed.[2] We have used copper and platinum-gold object holders, as well as nickel, and find the growth slightly less rapid on the platinum-gold and slightly more so on the copper disks. This gradation is in the order of the melting points of these metals, but it is not necessarily the case that this is the same as that of ease of sputtering. As a working hypothesis, however, we put forward the view that the observed increase in size of particles in good electrical contact with the grid arises from the deposition of charged metal and other particles ejected from the supporting holder by the electron beam. If it were from organic vapors, it should be observable in gold-shadowed viruses and similar specimens. In Watson's case it may be that contamination comes also from the carbon black itself under electron bombardment, and hence affects gold-shadowed specimens.

---

[*] I. C. I. Research Fellow.

[1] Watson, J. App. Phys. 18, 153 (1947).

[2] The original plates of Figs. 1 and 2 show the accretions of material to be of lower opacity to electrons than the zinc-oxide needles; this may, however, arise from loose agglomeration rather than from a smaller physical density.

# Ultrasonic Modulation of a Light Beam*

R. F. Humphreys, W. W. Watson, and D. L. Woernley**
*Sloane Laboratory, Yale University, New Haven, Connecticut*

June 23, 1947

THE use of a light beam as a carrier of intelligence has the obvious military advantage of high collimation, hence a low probability of detection by an enemy not in the direct line of the beam. If the wave-length of the light is outside the visible region, the chance of direct line-of-view detection by an enemy is still further reduced. However, there are three major limitations to the use of such a communication device: (a) attenuation of the carrier light beam by atmospheric absorption, (b) attenuation by the infra-red (or ultraviolet) filter, and (c) the difficulty in obtaining a satisfactory signal-to-noise ratio. Evidently, these limitations are somewhat interdependent. If, furthermore, it is desired that the equipment be portable, the limitation of power and space must be considered. This note is concerned with the performance of an experimental infra-red communication unit designed for the Army Signal Corps.

The most direct method of light modulation involves varying the intensity of the source of light in accord with the signal frequency. The very appreciable thermal inertia of light sources, however, permits only a relatively small signal-to-noise ratio. That an ultrasonic grating, formed by a traveling sound wave in a liquid medium, will serve as an almost inertia-less modulator has been pointed out by Jeffree[1] and Becker,[2] and the method has been employed successfully in the Scophony Television system.[1]

A beam of parallel light falls on a glass cell (Fig. 1) at the bottom of which is a 7-megacycle quartz crystal. The cell is filled with xylol and capped with a plug of glass wool to prevent the formation of standing sound waves. (The cell performs equally well as a modulator with either traveling or standing waves; however, it is not desirable to have a mixture of the two.) The waves traveling up the cell form moving regions of increased and decreased density, hence, are similar to a moving diffraction grating. The diffraction pattern which results when light is directed normally to the sound waves has an order spacing proportional to the wave-length of the crystal excitation (approximately 0.4° for the frequency used). The distribution of light intensity in the various orders depends on the amount of crystal excitation; the light in the zero order is nearly inversely proportional to the excitation of the crystal.[3] Thus, an ultrasonic grating corresponds somewhat to an optical grating whose groove shapes are variable at will. If the crystal excitation is modulated by a signal, the light received in the zero order will have the same modulation. It should be noted that orders higher than zero are likewise modulated, though 180° out of phase with the undeviated light; the advantage gained by using the zero order lies in its greater intensity.

The light source found most satisfactory was a 6-volt

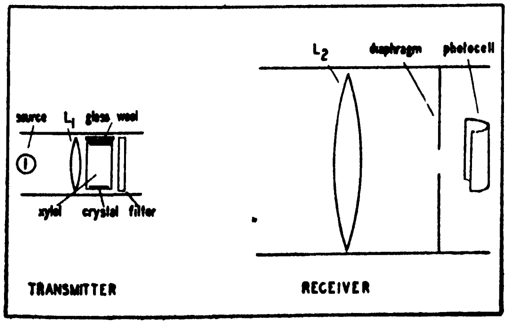


FIG. 1. Schematic diagram of the transmitter-receiver unit for modulating an infra-red light beam.

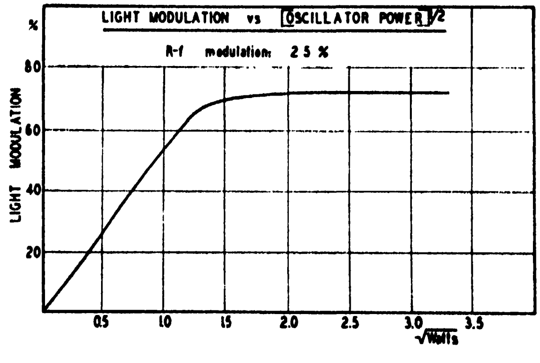


FIG. 2. Light modulation of the zero-order diffraction image as a function of sound intensity (assumed proportional to the voltage on the crystal).

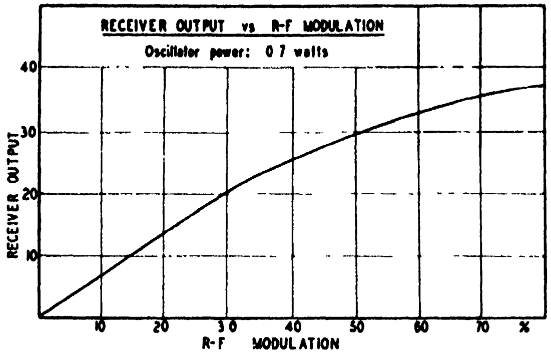


FIG. 3. The curve of audio-signal output as a function of percentage r f modulation indicates the approximate linearity of the ultrasonic modulator. At 70 percent modulation the deviation is only 2 db.

Mazda ribbon lamp, operated at 8 volts; at this latter voltage the infra-red energy was doubled, while the life of the lamp was not seriously shortened. An achromatic lens ($L_1$), of 60-mm diameter, formed parallel light which completely filled the cell. The infra-red filter (Corning "Heat Resistant") was made removable to allow daytime signaling. The collection lens ($L_2$) at the receiver should be as large as conveniently possible to obtain the maximum signal-to-noise ratio. An eight-inch diameter lens was used in the present work with a quarter-inch diaphragm placed at its principle focus, slightly in front of the photocell. A rifle telescope (magnification 4×) was mounted on the unit to aid initial contact between the two communicators.

A 7-megacycle x-cut quartz crystal rested on the stainless-steel base of the liquid cell; on the crystal's upper surface was evaporated copper and then a layer of gold. Contact with this electrode was made with a spring-bronze fork, soldered to the gold surface with Wood's metal to eliminate contact sparking. The crystal was driven at its fundamental frequency by a conventional electron-coupled oscillator with one stage of amplification; a 6V6GT tube served to plate-modulate the amplifier stage. The receiver contained an RCA 918 photo-cell with three 6Sj7 stages of audio-amplification; earphones were used in the output circuit. For the portable unit a 6-volt vibrator pack supplied the necessary power; the vibrator, transmitter, receiver, and optical system were contained in a single, well-shielded unit. The total power consumption was approximately 40 watts. The r-f power delivered to the crystal was of the order of one watt.

The linearity of the ultrasonic cell as a modulator is indicated in Fig. 2, where the sound intensity is taken to be proportional to the square root of the oscillator power. The degree of r-f modulation is held constant (25 percent). It is seen that from zero to 70 percent the modulation is reasonably linear. A curve of r-f modulation (constant r-f power) *versus* receiver output confirms this relation (Fig. 3). The shape of the curve conforms with that predicted by the theory of Raman and Nath.[4]

The most severe operating conditions occur, of course, in bright sunlight, where the signal-to-noise ratio is low. With the red filter removed, the limiting distance for satisfactory two-way conversation was found to be about 3000 yards. Night communication using the filter was very clear at 3000 yards and was still possible at 5000 yards. Beyond this the background noise became prohibitive. If the restriction of low power consumption is removed, a more intense source of light and a larger modulating crystal should easily double this maximum range.

---

* This paper is based on work done for the Signal Corps Engineering Laboratories under Contract W-1077-SC-1077 in 1942.

** Present address: Cornell Aeronautical Laboratory, Buffalo, New York.

[1] J. H. Jeffree, Television (London) **9**, 260 (1936).

[2] H. E. R. Becker, Zeits. Hochfrequenztech. **48**, 89 (1936).

[3] H. E. R. Becker, Ann. Physik (Leipzig) **25**, 373 (1936).

[4] C. V. Raman and N. S. N. Nath, Proc. Ind. Acad. Sci. **2**, 406, 413 (1935).

## Films, Resistant to Organic Solvents, for Use in the Electron Microscope

S. G. Ellis
*Department of Physics, University of Toronto, Canada*

June 28, 1947

FILMS of Alkathene,* insoluble in ether and suitable for specimen supports in the electron microscope have been made from hot xylene solutions of the plastic.

A suitable strength for the solution is 0.2 g of Alkathene grade 2 in 100 cc of xylene. It is convenient to use the solution in such a quantity (75 cc) that, when placed in a bottle with a wide neck, a clean glass microscope slide can be half-immersed in the bottle, and a ground glass stopper inserted. The bottle is heated in a water bath at 100°C until the solution is clear. The bottle may then be removed from the bath, the slide lifted from the solution and held vertically just above its surface for a few seconds and then removed from the bottle. The film of Alkathene left on the slide is suitably cut with a needle and then floated off on the clean surface of some hot distilled water at a

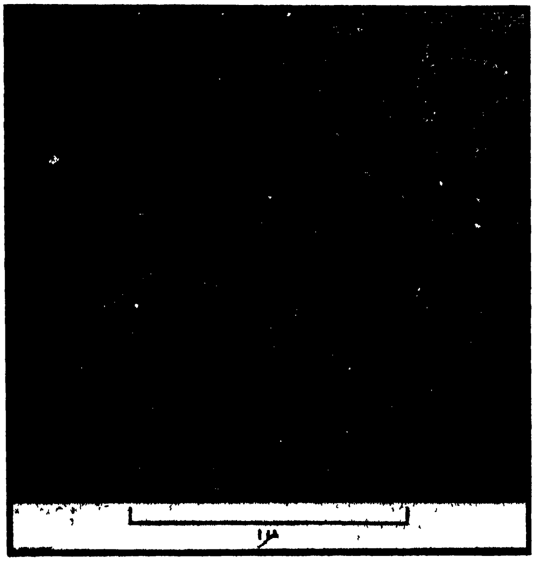


FIG. 1. Alkathene film shadow cast with chromium

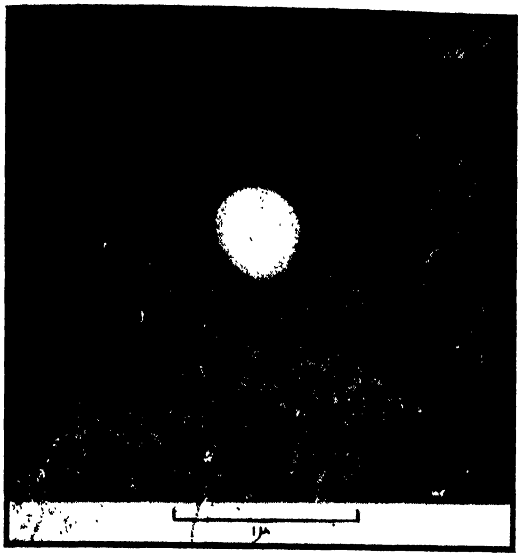


FIG. 2. Unidentified particle on Alkathene film shadow cast with chromium.

temperature above 50°C. A mesh can then be placed on the film and mesh lifted from the water and allowed to dry in the usual manner. When another microscope slide has been placed in the Alkathene solution it should be allowed to come to the temperature of the solution before being removed.

These Alkathene films, when examined in the electron microscope, show a characteristic structure (see Fig. 1). There are many applications in which the structure will not interfere with the observations since adequate contrast can be obtained in the micrographs (Fig. 2). The strength of these films, for a given transparency to the electron beam, is comparable with that of Formvar films. The films, when mounted on a mesh, have been flooded with either suspensions, both of soaps and carbons, without any apparent solvent action on the Alkathene.

In the studies made to date it has not been possible to prevent the formation of the characteristic structure in these films either by varying the strength of the xylene solution or by changing the conditions under which the xylene evaporates. Attempts to form a film by rapidly dipping and withdrawing a cold microscope slide into the hot xylene solution failed.

Similarly attempts to cast the films on the surface of hot water failed, since with the solvents used the solutions of the plastic spread in an irregular and unsatisfactory manner.

We are indebted to Mr. F. G. Rice of Canadian Industries Limited, for samples of Alkathene and for information on its use.

---

* "Alkathene" (registered trademark of I.C.I. in Canada and most of the important countries in the world) is a straight low polymer of ethylene formed at high pressure. Its properties are described in a series of papers by E. Hunter and W. G. Oakes, R. B. Richards and E. L. Midwinter, in British Plastics for March, April, and May, 1945. It is available from: Plastics Division, Canadian Industries, Ltd., P. O. Box 10, Montreal, Canada, for delivery in Canada only, and Plastics Division, Imperial Chemical Industries, Ltd., Welwyn Garden City, Hertfordshire, England.

# Note on the Reduction of Microphonics in Triodes[1]

VICTOR W. COHEN AND ABRAHAM BLOOM
*Ordnance Development Division, National Bureau of Standards, Washington 25, D. C.*

June 2, 1947

UNDER the above title, Mr. Waynick has considered the behavior of a simple triode with a vibrating grid and developed, as a consequence of the motion, a relationship between the microphonic output and the grid bias. This relationship indicates that under certain conditions the microphonic response can be appreciably reduced by a suitable choice of grid bias. The application of this effect to high gain amplifiers with fixed bias may be of importance. In the application of the theory however to the 6A3 tube, he has failed to consider two very important factors both of which pertain to this type of tube. These are: (a) The effect of symmetry in a plane symmetrical triode, and (b) the importance of motion of the cathode in filamentary tube types.

Consider an idealized symmetrical plane triode with the cathode in the center and two grid planes attached to a pair of side rods so that they vibrate with their separation fixed. Then as the grid-cathode distance *increases* on one side, it will *decrease* on the other. To a first-order approximation, therefore, the vibration of the grid side rods will result in no change in $I_p$. For a more accurate evaluation of the microphonic output, one must consider the second-order approximation. We have made such a calculation and find that, depending upon the spacings, there may or may not be a point between negative cut-off and zero grid bias at which the microphonic output is zero. A rough approximation of the second-order term is:

$$I-I_0=\Delta I=G_m\Delta d^2\left[\frac{4}{d^2}\left(E_g+\frac{K}{b}E_p\right)+\frac{(d-2b)}{bd}\,\frac{2}{b^2}KE_p\right.$$
$$\left.+\tfrac{1}{2}\frac{K^2E_p{}^2}{b^4}\left(E_g+\frac{K}{b}E_p\right)^{1}\right],$$

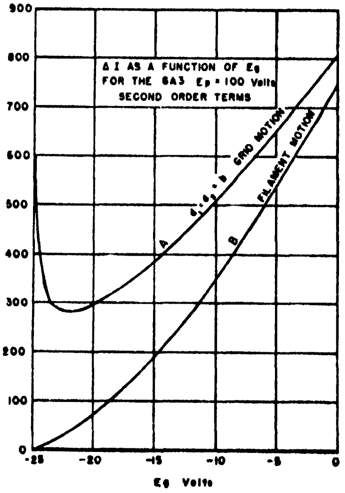


FIG. 1. The second-order term for the change in plate current due to (A) grid motion and (B) filament motion, as a function of grid bias. The curves are calculated for the 6A3, with 100 volts applied to plate.

where $G_m = 3C/2d^2[E_g+(KE_p/b)]^{\frac{1}{2}}$, where $d$ = grid-cathode spacing, $b$ = grid-plate spacing, $C = 2.33\times10^{-6}A$ amperes, $\Delta d$ = grid displacement, and $K = 1/2\pi n \log 1/2\pi np$.

The behavior of this second-order term for $\Delta I$ as a function of grid bias is shown in Fig. 1 curve A, where the term $\Delta Id^4/C(\Delta d)^2$ is plotted against $E_g$, for the special case of the 6A3 in which $d=b$, $u=4$, $E_p=100$ v.

This function goes to infinity at cut-off and has a minimum not far from cut-off.

Since in practice the tube is not perfect in its assembly, the symmetry may not be exact. The microphonic output due to grid vibration may, therefore, contain components of both first and second orders. The first-order term will obey a function similar to Waynick's but will have an amplitude dependent upon the degree of asymmetry.

For a single plane triode the first-order term of $\Delta I$ as a function of $E_g$ is shown in Fig. 2 curve A, in which $\Delta Id^3/C\Delta d$ is plotted against $E_g$ for the same case as Fig. 1A. If the vibration is harmonic, the first-order term of the output will be a sine wave which will reduce to zero amplitude at a particular bias as shown. The second-order term, however, since it depends upon the square of the grid displacement, will have a wave form of the second harmonic of the vibration frequency since $\sin^2\omega t = \frac{1}{2} - \frac{1}{2}\cos 2\omega t$.

While it is true that for indirectly heated cathodes the important cause of microphonics will be the grid vibration, in filamentary types the resonant vibration of the cathode is very important. We have subjected a 6A3 to vibration tests and find very distinct resonant vibration of the filament legs visible with the naked eye. Each leg had its own resonant frequency, most of these being near 800 cycles. The grid vibration was of smaller amplitude and was not visible in a binocular microscope of 36×, using stroboscopic illumination. It was detectable with a r-f capacity vibration pick-up.

The cathode of each section of the 6A3 is formed from a narrow ribbon, in the shape of a letter M. The plane of the ribbon in each leg is designed to be perpendicular to the idealized cathode plane. One would expect, therefore, that appreciable vibrations would occur only in the cathode plane. However, since in assembly the cathode ribbon may be warped slightly, there may be a component of vibration normal to the cathode plane. A rough calculation of the change in plate current due to filament motion gives to a second-order approximation:

$$I-I_0=\Delta I=-2\frac{C}{d_1^3}\left(E_g+\frac{K}{b_1}E_p\right)^{\frac{3}{2}}d_1+\frac{2C}{d_2^3}\left(E_g+\frac{KE_p}{b_2}\right)^{\frac{3}{2}}d_1$$
$$+3\frac{C}{d_1^4}\left(E_g+\frac{K}{b_1}E_p\right)^{\frac{3}{2}}(d_1)^2+\frac{3C}{d_2^4}\left(E_g+\frac{K}{b_2}E_p\right)^{\frac{3}{2}}(d_2)^2,$$

where the subscripts 1 and 2 refer to opposite sides of the triode.

Clearly all four terms of this expression are $\frac{3}{2}$ power curves and have minima only at cut-off, i.e., where $E_g = -K/bE_p$. The form of the second, and first-order terms are illustrated in Fig. 1 curve B and Fig. 2 curve B, respectively.

With the foregoing considerations in mind, one can explain some of Waynick's observations on the 6A3 which appear to depart from his theory.

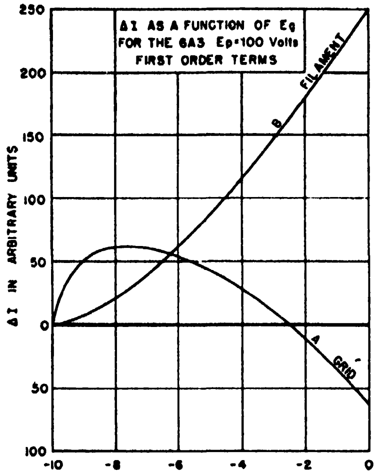


FIG. 2. The first-order term for the change in plate current due to motion of (A) grid and (B) filament as a function of grid bias. The curves are calculated for the 6A3, with 100 volts applied to plate.

1. Case 2 and 8 may be explained as due to resonances either of a filament leg or of the grid in a symmetrical section of the tube.
2. Case 6 was undoubtedly due in part to a second-order term giving rise to a second harmonic of the vibration frequency which did not pass through zero as the fundamental component did.
3. The residual output near the minima which appeared to contain the second harmonic of the driving frequency was probably due to second-order effects.

The apparent close agreement observed in some cases may have been due to a grid vibration in a highly asymmetrical section of the tube.

A more detailed discussion of calculations of the effects of vibration of triode elements as well as experimental methods and observations is contained in a manuscript now nearing completion.[2]

---

[1] Waynick, J. App. Phys. 18, 239 (1947).

[2] Paper delivered at Winter Convention of the I.R.E., March 1947.

---

## Comments on "Sound-Velocity Determination of Molecular Weights"

G. J. DIENES
*Development Laboratories, Bakelite Corporation, Bound Brook, New Jersey*

June 24, 1947

In a recent paper in this journal, A. Weissler, S. W. Fitzgerald, and I. Resnick[1] described a method of molecular-weight determination for liquid polymers based on sound-velocity, density, and refractive-index measurements. The essential relation for calculating molecular weights by their method is an empirical linear correlation between molar sound velocity and molar refraction.

It occurred to the writer that there is no necessity for making the sound-velocity measurements since the mo-

TABLE I.

| Glycol | Theoretical mole wt. | No. ave. mole wt. from specific refractivity | Deviation. % |
|---|---|---|---|
| Ethylene | 62.1 | 62.1 | 0.0 |
| Diethylene | 106.1 | 101.5 | −4.3 |
| Triethylene | 150.1 | 144.8 | +3.5 |
| Tetraethylene | 194.2 | 192.9 | +0.7 |
| Pentaethylene | 238.2 | 245.7 | −3.1 |
| Poly 200 | 200 ± 10 | 193.0 | 0.0 |
| Poly 300 | 300 ± 15 | 316.9 | +0.6 |
| Poly 400 | 400 ± 20 | 408.6 | 0.0 |
| Poly 600 | 600 ± 30 | 576.2 | 0.0 |
| | | Ave: | 1.4% |

lecular weights are easily calculable from molar refractions, as pointed out recently by E. Warrick.[2] Calculation of molecular weights from molar refractions is based on the well-accepted concept of group additivity and the additional step of an empirical correlation between molar velocity and molar refraction is eliminated. Since the molar-refraction method is physically analogous to a chemical end-group determination, the resulting molecular weight, in the case of a heterogeneous polymer, will evidently be a number average. This can also be readily shown mathematically using additivity of molar refractions on the basis of mole fractions.

In the system of polyethylene glycols, treated by the above authors, the average experimental molar-refraction increment for each $-CH_2-CH_2-O-$ structural unit, based on the first five pure members of the series, is 10.97. From the general formula for the polyethylene glycols, $HO-(CH_2-CH_2-O-)_n-H$, the molecular weight, $M$, of any member may be written

$$M = 44.05i + 62.07, \qquad (1)$$

where

$i$ = number of $CH_2-CH_2-O$ units in compound, 44.05 = molecular weight of $CH_2-CH_2-O$ unit, and 62.07 = molecular weight of ethylene glycol.

Similarly, the molar refraction, on the basis of additivity of groups, may be expressed

$$R = (n^2-1/n^2+2)M/d = rM = 10.97i + 14.47,$$

or

$$r(44.05i + 62.07) = 10.97i + 14.47, \qquad (2)$$

where

$r = (n^2-1/n^2+2)1/d$ = specific refractivity of compound, and 14.47 = molar refraction of ethylene glycol.

Equation (2) is easily solved for $i$, and substitution of $i$ in Eq. (1) gives the molecular weight. From the density and refractive-index data reported by Weissler, Fitzgerald, and Resnick, the molecular weights given in Table I were calculated.

The agreement between theoretical molecular weight and molecular weight calculated from the experimental specific refractivities is actually better than that obtained by the molar-velocity method (1.4 percent *vs.* 2.0 percent). Since the molar-refraction method depends essentially on differences between the molar refractivity of the chain and the end-groups, it is evidently only applicable to compounds of relatively low degree of polymerization. This same limitation, however, applies to the molar-velocity method. For low molecular-weight liquid polymers the molar-refraction method is very useful and convenient as the necessary density and refractive-index measurements are easily made to a high degree of accuracy.

---

[1] A. Weissler, S. W. Fitzgerald, and I. Resnick, J. App. Phys. **18**, 434 (1947).

[2] E. Warrick, J. Am. Chem. Soc. **68**, 2455 (1946).

# Preliminary Measurements Relative to the Onset of Thermal Convection Currents in Unconsolidated Sands*

H. L. MORRISON
*North Carolina State College of Agriculture and Engineering, Raleigh, North Carolina*

July 10, 1947

TO assess a recently advanced theory[1] for the minimum temperature gradient necessary to cause the onset of convection in a viscous liquid in a porous medium, a series of appropriate measurements has been initiated using several different liquids separately in unconsolidated sands of known permeability. These measurements are reported here in a preliminary form because of their interesting bearing upon the theory itself and upon its application to geophysical problems. The measurements were made on mixtures (depth 10 cm) of liquid and sand in a 400-cc beaker (diameter 7.0 cm) which was heated from below; temperatures were observed at various depths in the mixture, as functions of time; and microscopic observations were made of the onset of motion in small suspended solid particles (impurities) which were visible near the wall of the beaker. Precautions were taken to assure the horizontal uniformity of temperature (maximum deviation about ¼°C to ½°C) throughout the mixture; and subsidiary measurements of upward motions of the liquids, caused by diffusion, were made using suitable soluble dyes so as to assure that this possible disturbing effect was negligible. A further subsidiary series of measurements was made using a beaker of 15.0-cm diameter, to exhibit the effects of vessel radius; there were no appreciable effects on the temperature gradient required for onset of convection. Typical observations are recorded in Table I. In Table I, the data in columns 4 to 6 are cast in the form used in reference 1, so that $z$ column 4+column $6 = z + \theta$; the column 3 refers to the gradient at $z$ (distance into mixture

TABLE I.

| | | Temperatures and gradients at onset of convection | | | |
|---|---|---|---|---|---|
| Liquid | Permeability of sand (d'Arcy's) | Local gradient (°C/cm) estimated | β-gradient (°C/cm) | Temperature at $z = 0$ (°C) | Approximate θ temperature function (°C) |
| Glycerol | 95 | −80 | −8.4 | 125 | $-47\frac{1}{2}\sin(\pi z/10)$ |
| $CCl_4$ | 127 | −0.8 | −0.8 | 31½ | 0 |
| $H_2O$ | 127 | −6.0 | −6.0 | 57 | 0 |

from bottom) = 2 cm, the usual point of location of the microscope.

In attempting to compare the data in Table I with the theory,[1] one meets definite difficulties. Thus, the $\theta$-function for the $CCl_4$ and $H_2O$ data, does not accord with the sinusoidal one assumed in developing the theory, though this may not be too significant, for the theory might use a more general series of sine functions which could correlate with the zero used in the measurements. Again, the theory does not contemplate a variation of viscosity with temperature, a situation which is far from acceptable for the liquids used in these measurements. Finally, the finite horizontal extent of the mixture in these measurements implies, at least in principle, boundary conditions which the theory does not contemplate. Nevertheless, any attempted application of the Eq. (26) in reference 1 to these experiments appears to indicate that for the glycerol test, the calculated gradient is at least on order of magnitude in excess of the observed, and for the other two, the calculated gradient is slightly in excess of the observed. Apart then, from the possibility that it is not correct to identify the onset of motion in suspended matter in the present series of measurements with the onset of convection, it can only be concluded at the moment that the theory[1] predicts excessive values for the minimum temperature gradient but that it deserves such extensions as to render it applicable to materials, dimensions, and $\theta$-function having actual, rather than ideal, characters.

This work was suggested to the author by Dr. F. T. Rogers, Jr., to whom, with Mr. C. W. Horton, the author is also indebted for continuing conversations and correspondence relating to it. Further measurements are in progress, and will be reported when and if it is possible to extend the theory suitably.

---

* An earlier report on this work was presented at the April 1947 meeting of the Southeastern Section of the American Physical Society at Salisbury, North Carolina. See Phys. Rev. 71, 834 (1947).

[1] C. W. Horton and F. T. Rogers, Jr., J. App. Phys. 16, 367 (1945).

---

# Attachments for Aligning a Console Electron Microscope

HENRY C. FROULA
*Armour Research Foundation of Illinois Institute of Technology, Chicago, Illinois*

July 8, 1947

THE alignment of the electron gun and anode aperture in our RCA console electron microscope was greatly simplified through the use of the two devices described below.

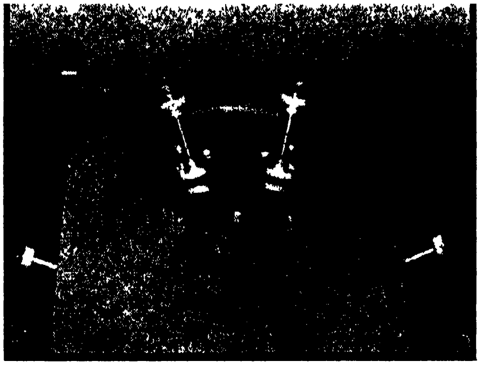


FIG. 1. Rear view of console electron microscope, showing external controls for alignment of electron gun and anode aperture.

The first of these two devices provides for the alignment of the electron gun. Three arms mounted 120° apart on the front rim of the gun extend forward to overhang the fixed column by about one-half inch. One of these arms (on the underside of the gun) is a flat spring whose free end presses upward against the column, tending to lower the gun. Each of the other two arms has a tapped hole through which a screw is driven against the fixed column. Manipulation of the screws provides a two-dimensional adjustment of the gun position. The two screws can be turned externally by means of the removable extension shafts which project through the sides of the hood as shown in Fig. 1.

The second of the two devices provides for the external adjustment of the anode aperture. The regular cover of the specimen chamber is replaced temporarily by one having two sealed passages holding a pair of metal rods. These rods can be seen projecting upward in the figure. Their lower ends are articulated with the two screws which adjust the anode aperture. The vacuum seals through which the rods pass, permit a liberal amount of movement of the rods so that adjustments can be made during actual operation of the microscope.[1]

Both of these devices have had considerable use in our laboratory and have given very satisfactory service in establishing and maintaining suitable alignment of the microscope.

---

[1] The seals are of the type described by D. B. Cowie, Rev. Sci. Inst. 15, 46 (1944).

# New Books

## Radium Therapy; Its Physical Aspects.

By C. W. Wilson. Pp. 224, Figs. 97, 30 tables, and 4 mathematical appendices. Chapman and Hall, Ltd., London, 1945. Price $6.00 plus postage.

This book supplies an excellent practical survey of the physical aspects of radium therapy, more complete and detailed than anything hitherto available. The basic theories of radioactivity are passed over rather briefly, but the interaction of high energy radiations with matter is discussed in detail. The development of methods and theory of dosimetry receives considerable attention. Detailed descriptions are given of the air-wall ionization chamber, and of various practical vacuum-tube dosimeters, with instructions for their use.

A large part of the book is devoted to dosage calculations, with many examples. For surface therapy, the Paterson and Parker charts are followed. In intracavitary treatments, various systems for treatment of cancer of the cervix are compared in detail. For interstitial dosage calculations, the Memorial Hospital and Paterson and Parker techniques are discussed, with numerous examples for the latter. A chapter devoted to dosage in teleradium therapy will find few readers in America, since this method is now little used here. The matter of protection is adequately covered.

One might wish that more attention had been devoted to the matter of integral dose, or total energy absorbed by the patient, since this had been more thoroughly studied in England than in the United States.

Each chapter contains a good list of references, specific and general. The appendix consists of useful mathematical tables.

Radiologists have long felt a need for such a book; there has been no satisfactory text covering this field. It should be welcomed by students, teachers, and all practicing radiologists and radium therapists.

Edith H. Quimby
*Columbia University*

## Introduction to Atomic Physics

By Henry Semat. Revised Edition. Pp. 412+xi, Figs. 169, Tables 27, 23×15.5 cm. Rinehart and Company, Inc., New York, 1946. Price $4.50.

The first edition of Semat, appearing in 1939, at once became a standard undergraduate text on atomic physics and won a place of respect on the shelves and desks of physicists and chemists of experience. This second edition is substantially the earlier Semat, though it presents an altered appearance. The page size, type, paper, and some of the subdivision of text into chapters have been changed along with the corporate title of the publisher. The effect of the new format is one of increased elegance and modernity, but not all users will approve the change to paper of half the former thickness.

Getting down to subject matter, one finds that a large number of small, first-edition errors have been corrected, data have been brought up to date, several diagrams have been redrawn, and new paragraphs inserted to clarify or supplement previous treatments. This piecemeal modernization has been accomplished smoothly without serious impairment of the unity of the book. Some new problems have been constructed and some old ones improved. The statement that 94 elements are found in nature is not yet correct.

The book begins and ends with nuclear energy, the frontispiece comprising photographs of the first atomic bomb explosion, and the final chapter being devoted entirely to a discussion of nuclear energy. This chapter, the only completely new one in the book, starts with the stellar energy reactions of Bethe and proceeds through a careful history of nuclear fission researches to nuclear chain reactions, fast and slow. Some of the latter discussion is taken sentence for sentence from the Smythe report, without quotation marks, but with a preceding statement that the discussion is based on the report. The reviewer dislikes and disbelieves statements that mass is converted into energy, or even "other forms of energy." This common stereotype will do no harm to physicists but may interfere with students' understanding of the equivalence equation since it suggests that after the event the total mass is less than it was before.

Other topics which make their first appearance in this edition are the measurements of nuclear magnetic moments, the discovery and production of new elements, electron optics, and the design and use of the betatron. Semat's first edition was distinguished by an unpretentious clarity of statement and a consistent awareness of the capacity and limitations of the chosen audience. These characteristics are preserved in the current modifications and extensions. It is a sad circumstance that enables a book on atomic physics to be brought up to date after seven years with no more enlargement than Semat has found necessary. It is to be hoped that the third edition will require the inclusion of much more new material.

Paul Kirkpatrick
*Stanford University*

# New Booklets

The National Research Council calls attention to *Mathematical Tables and Other Aids to Computation*, a quarterly journal published by the Council at 2101 Constitution Avenue, Washington 25, D. C. It is edited by Raymond C. Archibald and Derrick H. Lehmer and is an international publication which serves as an information center of published and unpublished mathematical tables and material dealing with other aids to computation. Because of the continuous pagination in each volume and a system of constant cross references, new subscribers will find it desirable to begin their subscriptions with the first issue of the year. Subscription is $4 per year; single issues $1.25.

Jarrell-Ash Company, 165 Newbury Street, Boston 16, Massachusetts, announce their new 20-page catalog *I-3* describing their stigmatic grating spectrographs and accessories.

*Interchemical Review*, 432 West 45th Street, New York 19, New York, features in its Spring 1947 issue articles on the application of filters in photography, methods of printing on vinyl film, and the wartime development in Britain of a non-metallic conductive coating for paper.

The most recent issue (Volume 9, Number 2) of *Philips Technical Review*, published by Philips Laboratories in Holland, contains the following articles, according to North American Philips Company, Inc., 100 East 42 Street, New York 17, New York: "A new electron microscope with continuously variable magnification," by J. B. le Poole; "Electronic conductivity of non-metallic materials," by E. J. W. Verwey; "Installations for improved broadcast reception," by P. Cornelius and J. van Slooten.

RCA Victor Division, Camden, New Jersey, has issued a 16-page booklet, for distribution to consumers, entitled *What's Your Television I.Q.?*

Dover Publications, 1780 Broadway, New York 19, New York, has published a new 1947 catalog of scientific books published by it. 38 pages.

Burrell Technical Supply Company, 1942 Fifth Avenue, Pittsburgh 19, Pennsylvania, has issued an *Announcer* of scientific equipment. It is published quarterly and describes various laboratory instruments and equipment. Each edition also contains a dissertation on the history of some instrument or process now used in industrial research.

Cannon Electric Development Company, 3209 Humboldt Street, Los Angeles 31, California, has issued *Cannon Plugs*, a 78-page booklet containing many photographs and describing the uses of its products in industry.

Microfilm Service, a new affiliate of the American Council on Public Affairs, located at 2153 Florida Avenue, Washington 8, D. C., is publishing through the microfilm medium scholarly manuscripts on specialized subjects, documents of historical value, and out-of-print studies.

The Editorial Service Bureau of Kodak, 343 State Street, Rochester 4, New York, has issued the first edition of *Shorts*, a 4-page digest of new uses of photography.

Foster D. Snell, Inc., 29 West 15th Street, New York 11, New York, has released a 1947 edition of its brochure under the heading *The Chemical Consultant and Your Business*. Available without charge.

Distillation Products, Inc., 755 Ridge Road West, Rochester 13, New York, has published a 16-page booklet entitled *Information on High Vacuum Distillation*.

Fischer and Porter Company, Hatboro, Pennsylvania, has released a new, larger catalog section No. 95-D describing the company's line of precision bore glass specialties. Address Department 6F-D in requesting a copy.

# Here and There

### New Appointments

Cleveland Norcross, formerly executive secretary of the Office of Scientific Research and Development, is now assistant director of the American Institute of Physics.

Captain Robert D. Conrad, U. S. Navy retired, has been named assistant director of Brookhaven National Laboratory for atomic research.

Effective September 1, T. Keith Glennan, recently an executive of Ansco Division of General Aniline and Film Corporation and wartime director of the U. S. Navy Underwater Sound Laboratory, became president of Case Institute of Technology, Cleveland, Ohio, succeeding William E. Wickenden, who has retired.

*Science* reports that Henry N. Russell, professor of astronomy and director of the observatory, Princeton University, retired July 1. Lyman Spitzer, Jr., professor of astrophysics, Yale University, has been appointed director of the observatory, and Martin Schwarzschild, Columbia University, has become professor of astronomy.

William E. Good has been granted a year's leave of absence from the Westinghouse Research Laboratories to aid in atomic energy investigation at the Argonne National Laboratory near Chicago.

Louis N. Ridenour, formerly assistant professor of physics, University of Pennsylvania, has been appointed dean of the Graduate School, University of Illinois.

### New Casting Resin

P. J. Franklin and M. Weinberg of the National Bureau of Standards issued recently a report on a casting resin developed by the Bureau. Their report says, "During the war, exacting mechanical and electrical stability requirements of special electronic applications—such as the radio proximity fuze—necessitated the potting of the circuit components. Because of the high impedance of the circuits involved, the electrical loss factor or the dissipation of the available energy in the casting or 'potting' compound became a matter of major importance. An extensive investigation of various available casting resins at the National Bureau of Standards revealed none suitable for the particular applications because of the high loss factor. The few materials with adequate electrical properties were deficient mechanically. In addition, the applications required a resin of such viscosity that the potting compound would encompass all the circuit elements when poured into a container housing the electronic device, and then harden to a rigid solid without adversely affecting circuit operation.

"Many casting resins have been developed, but few have the vital electrical properties for proper operation of high-impedance, high-frequency equipment. As a result of experience gained in the fuze-development program at the National Bureau of Standards, a new casting resin embodying the required properties of mechanical and electrical stability, was formulated through a systematic variation of resin constituents. Various applications at the Bureau indicate that, with slight modifications to suit the intended use, the new potting compound—known as NBS Casting Resin—can be readily employed in many high-frequency devices requiring such mechanical-electrical insulation.

"The most important properties specifically desired of a casting resin when utilized at high frequencies in high-impedance circuits are low power factor, low dielectric constant, short polymerization period at low temperature and atmospheric pressure, high impact strength, small volume shrinkage on polymerization, dimensional and electrical stability, and low moisture absorption. The NBS Casting Resin not only meets these requirements but fulfills the additional requirements of low viscosity and low surface tension, and hence may be poured through small openings."

### New Research Laboratory in Texas

In April a new research laboratory was officially opened on a 3000-acre strip eight miles west of San Antonio, Texas. The project, made possible by the gift of a valuable oil property by Tom B. Slick, San Antonio oil and ranch man, consists of two separate but closely associated units: the Foundation of Applied Research, utilizing the Essar Ranch as part of its laboratory, and the Institute of Industrial Research, including a development service for inventors. While the Foundation of Applied Research was incorporated by its trustees as a non-profit undertaking, it has been designed to make its own way financially, after a development period, from sales of Essar Ranch's blooded cattle and from earnings by the Institute of Industrial Research through fees and patent proceeds.

Dr. Gregory Pincus, head of the Worcester Foundation for Experimental Biology at Worcester, Massachusetts, is selecting the research men and scientists who will staff the Foundation and the Institute.

Objectives of the projects have been outlined by Mr. Slick as follows: "The only limitation to the scope of work to be undertaken by the Foundation is that it has in view a distinctly practical application. The value of pure scientific study is fully realized, but it is felt that the Foundation's greatest contribution can be realized by concentration on the field lying between pure scientific study as carried out by universities and similar institutions, and completely industrial research of limited range as conducted by the average industrial corporation."

The Institute of Industrial Research, under separate management from the Foundation, will undertake, on a fee basis for industrial corporations, programs similar to the services rendered by the Mellon, Armour, and Midwest Institutes, and in addition will provide inventors with a complete processing service should their projects be determined worthy of development.

### New Computer at California Institute of Technology

*Science* reports that an electric analog computer will be part of the new analysis laboratory now being established at the California Institute of Technology. Gilbert D. McCann, Associate Professor of Electrical Engineering at the Institute, is setting up the laboratory.

The new calculator, which will be made available to all interested industrial and engineering organizations as well as being used for research at the Institute, is expected to be completed by August. Weighing approximately 33,000 pounds, the machine, including its control desk, will occupy some 1000 square feet of floor space. The device was designed for the solution of complex mathematical problems in the field of algebra and calculus requiring an accuracy of the order of 1 percent. It is particularly suited to solving engineering problems, including those of electrical circuits; applied mechanics; self-induced mechanical oscillations, such as airplane structure flutter, car shimmy, rail car stability; problems in thermodynamics, such as a general analysis of heat flow problems; speed controllers, voltage regulators; and angle position controllers, such as airplane stabilizers, guided missiles, gyro pilots, etc.

### Physical Society Acoustics Group

The Acoustics Group of the Physical Society was formed at an inaugural meeting at the Royal Institute of British Architects in February 1947. Dr. Alex Wood gave an address entitled "The Contribution of Acoustical Science to Allied Studies." The meeting was attended by some 170 persons. The principal object of the Group is to provide an opportunity for the varied types of workers engaged on acoustical problems to meet and discuss the scientific and technical implications of their work. Membership is open both to members and also to non-members of the Physical Society. For further particulars inquire of the Joint Honorary Secretaries of the Acoustics Group, The Physical Society, 1 Lowther Gardens, Prince Consort Road, London S. W. 7.

### Sound Apparatus Company

The Sound Apparatus Company announces the opening of its production plant in Stirling, New Jersey. The company's research laboratories are in Millington, New Jersey, and the main offices are at 233 Broadway, New York City.

### High Vacuum Symposium

National Research Corporation, in cooperation with the Division of Industrial and Engineering Chemistry, American Chemical Society, has announced plans for a High Vacuum Symposium to be held in Cambridge, Massachusetts, October 30 and 31. Secretary for the event is Stanley Heck, National Research Corporation, Boston 42, Massachusetts.

### Kodak Fellowships

Sixteen educational institutions have been offered one or more of twenty-two Eastman Kodak fellowships in chemistry, physics, engineering, and business administration for the school year 1947–48. In addition six similar fellowships are to be sponsored by Tennessee Eastman Corporation, Kodak subsidiary.

### National Electronics Conference

A National Electronics Conference will be held at the Edgewater Beach Hotel, Chicago, Illinois, November 3–5, 1947. Dr. George D. Stoddard, new president of the University of Illinois, will deliver the keynote address. The complete program for the Conference includes twenty technical sessions with a total of about fifty papers.

### Taylor Elected Radiology Fellow

Lauriston S. Taylor, chief of the x-ray section at the National Bureau of Standards, has been elected an associate fellow in the American College of Radiology.

### Institute of Navigation Officers

The institute of Navigation has elected the following national officers for the academic year 1947–48: president, Commodore G. G. McLintock, USMS, Special Assistant to the Commandant, U. S. Maritime Service; vice president, Captain P. V. H. Weems, USN (*Retired*), Annapolis, Maryland; technical adviser, Paul Rosenberg, president of Paul Rosenberg Associates, New York City; executive secretary, Samuel Herrick, professor of astronomy, University of California at Los Angeles.

### Chicago Physics Club Officers

The following are the new officers of the Physics Club of Chicago, elected May 1947: president, W. F. Einbecker; vice president, H. A. Carter; treasurer, J. W. Juvinall; secretary, H. C. Froula.

### Conference on Administration of Research

A conference on "Administration of Research" will be held at The Pennsylvania State College on Monday and Tuesday, October 6 and 7. The main purpose of the conference is to provide free discussion on the organization, administration, and operation of research laboratories.

Discussion will be led by Prof. Philip M. Morse, Brookhaven National Laboratories, and Dr. Jesse E. Hobson, Armour Research Foundation. The following men will be the principal speakers:

Dr. Maurice Holland, New York, Consultant
Dr. R. L. Jones, Bell Telephone Laboratories
Dr. G. H. Young, Mellon Institute of Industrial Research
Dr. Dwight E. Gray, Applied Physics Laboratory
Dr. L. Warrington Chubb, Westinghouse Electric Corporation
Dr. Edw. U. Condon, National Bureau of Standards
Hon. James H. Duff, Governor, Commonwealth of Pennsylvania
Commodore Henry A. Schade, Director, Naval Research Laboratory
Dr. Paul D. Foote and Dr. Blaine B. Westcott, Gulf Research Laboratories
Dr. Albert W. Hull, General Electric Company

For further information and a copy of the program, address Dr. Eric A. Walker, Head, Electrical Engineering Department, The Pennsylvania State College, State College, Pennsylvania.

### Calendar of Meetings

| | |
|---|---|
| **October** | |
| 2–4 | Society of Automotive Engineers, Los Angeles, California |
| 15–18 | American Society for Civil Engineers, New Orleans, Louisiana |
| 20–24 | American Society for Metals, Chicago, Illinois |
| 23–25 | Optical Society of America, Cincinnati, Ohio |
| 30–31 | National Research Corporation and American Chemical Society, Cambridge, Massachusetts (High Vacuum Symposium) |
| **November** | |
| 3–5 | National Electronics Conference, Chicago, Illinois |
| 3–7 | American Institute of Electrical Engineers, Chicago, Illinois |
| 6–7 | Society of Automotive Engineers, Tulsa, Oklahoma (Fuels and Lubricants Meeting) |
| 7–8 | Conference on X-Ray and Electron Diffraction, Mellon Institute of Industrial Research, Pittsburgh, Pennsylvania |
| 17–19 | Institute of Radio Engineers and the Radio Manufacturers Association, Engineering Department, Rochester, New York |
| 28–29 | American Physical Society, Houston, Texas |
| **December** | |
| 1–3 | Society of Automotive Engineers, Kansas City, Missouri (Air Transport Engineering Meeting) |
| 1–5 | American Society of Mechanical Engineers, New York, New York |
| 26–31 | American Association for the Advancement of Science, Chicago, Illinois |
| 29–31 | American Physical Society, Chicago, Illinois |

Journal
of
Applied Physics

Volume 18, Number 10 October, 1947

# Fringing Flux Corrections for Magnetic Focusing Devices

NORMAN D. COGGESHALL
*Gulf Research and Development Company, Pittsburgh, Pennsylvania*

(Received March 27, 1947)

When a charged particle approaches a region of uniform magnetic field its trajectory is affected by the fringing field. Compensation for the effect of the fringing field may in some cases be achieved by the use of virtual field boundaries which are outwardly displaced from the actual boundaries. A definition of the virtual boundary in terms of the effect of the fringing field upon the trajectories is formulated. A procedure for calculating the displacement *d* of the virtual from the actual field boundary is given using the methods previously developed for calculating ion trajectories in non-uniform magnetic fields. Using this procedure several values of the displacement for various magnet dimensions and conditions have been calculated. A method of altering the geometry of sector-type mass spectrometers to correct for the effects of the fringing fields is proposed. It is shown that the angular and spatial separations of the individual ion trajectories in an ion beam of small angular spread as it approaches a uniform field region through a fringing field are approximately the same as for the ideal case wherein the magnet field changes discontinuously at the boundary.

IT is well known that the magnetic field of a magnet does not discontinuously go from its maximum value to zero as the boundary of the region between the pole faces is passed. Rather the field falls off gradually in conformation with the laws of potential theory, and the result is an appreciable amount of fringing flux for some distance away from the pole-face boundaries.

In the sector-type mass spectrometer, schematically illustrated in Fig. 1, the ions proceed from the ion source through a magnetic-field-free region to the magnetic analyzer. The magnetic analyzer consists of a uniform magnetic field perpendicular to the plane of the figure and with the shape of a truncated sector. After passing through the magnetic analyzer the ions again pass through a magnetic-field-free region on their way to the ion collector. In the derivation[1] of the focusing properties of this system, it is assumed that the magnetic field region has sharp boundaries such as delineated in Fig. 1. Actually no magnet provides sharp boundaries to the magnetic field region, and in calculating the geometrical arrangement of ion source, magnetic analyzer, and ion collector, the effects of fringing magnetic field must be taken into account. One means of doing this is to use virtual boundaries of the sector-shaped field region which are displaced outwards from the actual boundaries. In this method it is assumed that fringing flux corrections can be effectively made by such a simple displacement of field boundaries. It is the purpose of this article to provide an analytical method for the determination of *d*, the displacement of the

---

[1] W. E. Stephens, Phys. Rev. 45, 513 (1934).

virtual field boundaries from the actual field boundaries. This displacement is considered under various conditions and several numerical cases evaluated. It is a further purpose of this article to consider the effectiveness of such a procedure as the use of virtual field boundaries for fringing flux corrections and to provide an alternative procedure which is believed to be more effective.

If we consider the magnetic field on the median plane, i.e., the plane parallel to and midway between the pole faces, the flux on it may to a reasonable approximation be regarded as a function of only one variable, the perpendicular distance from the boundary of the uniform field region. Thus the methods previously developed[2] for the calculation of ion paths in non-uniform magnetic fields may be used to determine analytically the trajectories of the central ray and others between the ion source and their penetration into the uniform field region.

In the past there have been indicated approximate methods of correcting for stray fields,[3–5] but they do not allow the exact calculations of the trajectories as will be described below.

In reference 2 it was shown that the exact determination of ion paths could be readily achieved for such cases where the ions move in a plane through a non-uniform magnetic field which is a function of only one coordinate and is everywhere perpendicular to this coordinate and to the orthogonal coordinate in the same plane. This was shown to be true for both Cartesian and polar coordinate systems, the magnetic field being a function of the radial distance in the latter. For the case at hand we may use the distance from the uniform field region and on a line normal to its boundary as the $x$ coordinate in a Cartesian system and apply these methods.

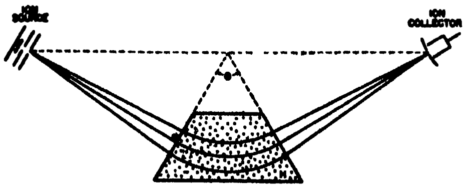


FIG. 1. Focusing scheme used in sector-type spectrometer.

In reference 2 it was shown that the ion trajectory is determined from the differential equation:

$$dy/dx = \pm f/(1-f^2)^{\frac{1}{2}}, \qquad (1)$$

where $f=f(x)$ and is defined as:

$$f = (e/vmc)\int H(x)dx,$$

and $v$ = velocity of ion in cm/sec., $c$ = velocity of light, $e$ = electrical charge of the ion in e.s.u., and $m$ the mass of the ion in grams. For convenience in the present applications we define a function $h(x)$ as follows:

$$H(x) = H_0 h(x).$$

The function $h(x)$ then specifies the ratio of the magnetic field at any point to the maximum field $H_0$ in the uniform field region. With this we may rewrite (1) as:

$$dy/dx = \pm f/(r^2-f^2)^{\frac{1}{2}}, \qquad (2)$$

where

$$f = f(x) = \int h(x)dx \quad \text{and} \quad r = (vmc/H_0 e).$$

The quantity $r$ is the radius of curvature of the ion path in the uniform field $H_0$.

It is to be noticed that the right-hand portion of Eq. (2) is a function of $x$ only. Therefore, an integration, either numerical or analytical, may be made to determine $y$ as a function of $x$ and thus obtain the trajectory. If the function $h(x)$ is known only experimentally then $f(x)$ may be determined by numerical integration and the resulting values used in a numerical integration of Eq. (2).

The application of the foregoing equations may be seen in Fig. 2. Here is plotted the trajectory of an ion as it passes from the ion source to and through a uniform magnetic field. Although the ion source is so oriented that its axis is perpendicular to the boundary of the magnetic field region, the fringing flux causes the central ray to actually enter the uniform field region with an angle other than 90° and at a

---

[2] N. D. Coggeshall and M. Muskat, Phys. Rev. 66, 187 (1944).

[3] A. J. Dempster, Phys. Rev. 11, 316 (1918).

[4] R. Herzog, Zeits. f. Physik 89, 447 (1934).

[5] J. Mattauch and R. Herzog, Zeits. f. Physik 89, 786 (1934).

point considerably displaced from the entrance point of the axis. The net effect of the fringing field is to so determine the trajectory that as the ion passes through the uniform field region the center of curvature is a distance $d$ normal to the pole-face boundary. This is in contrast to the ideal case wherein the magnetic field would be discontinuous at the boundary and the center of curvature would be on the boundary. We may therefore define a virtual boundary as one parallel to the actual boundary and displaced from it so as to include as a point the center of curvature for that part of the trajectory in the uniform field region. We shall consider $d$ at present only for the case wherein the axis of the ion source is perpendicular to the pole-face boundaries. As will be discussed in more detail later, $d$ depends upon the distance of the ion source from the uniform field region and upon the angle of emergence of the ions from the source. In Fig. 2 the virtual boundary is represented by the dashed line which contains the center of curvature as a point. It is useful to know the displacement $d$, although as we shall see later these virtual boundaries do not allow the use of the ideal geometrical arrangements considered in reference 1.

The determination of the displacement $d$ is readily made from Eq. (2). We note that $dy/dx = \infty$ for a value $\bar{x}$ such that $f(\bar{x}) = \pm r$. Suppose the actual pole-face boundary is taken as the line $x=0$, then since in the $x$-coordinate system the center of curvature will be located $r$ units from $\bar{x}$, it follows that:

$$d = (r - |\bar{x}|). \qquad (3)$$

The absolute value is used here since $\bar{x}$ may have a negative value, depending on manner of using the coordinate system. To obtain $\bar{x}$ we must use the relationship

$$\int_{x_0}^{\bar{x}} h(x)dx = \pm r, \qquad (4)$$

where $x_0$ is the value of the $x$ coordinate at the ion source exit slit.

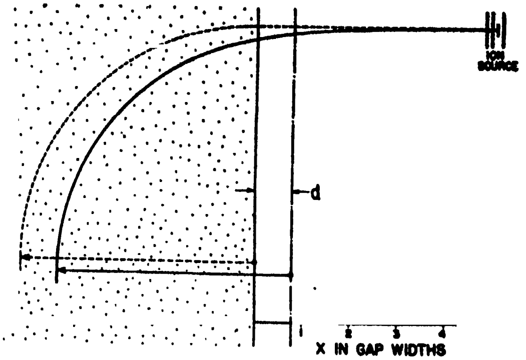


FIG. 2. Comparison between ion trajectories for the ideal case and for the actual case wherein the fringing flux is effective.

Thus the problem of determining $d$ reduces to obtaining the function $h(x)$ and to carrying out the integration of Eq. (4). Actually for any particular magnet $h(x)$ should be determined experimentally. We may, however, get some idea of the order of magnitude of $d$ and its change with variations of magnets by using the Schwarz-Christoffel transformation[6] to obtain $h(x)$. Figure 3 shows the type of magnet pole considered and the differential equation used for the conformal transformation. Here the magnet poles are considered to extend to $\infty$ in both directions perpendicular to the paper. They are in the form of jaws with the outer surface of each jaw being $n$ times further from the median plane than the inner surface. For simplicity the transformation is made for the case shown by the cross-hatching in Fig. 3, i.e., one magnet pole as one surface of constant magnetic potential and the median plane as the other. The differential equation shown transforms the edge of the upper magnet pole onto the left-hand half of the real axis in the $\omega$ plane and transforms the edge of the other region of constant magnetic potential onto the right-hand half of the real axis. In the differential equation $T$ is the usual orientation factor necessary for the transformation. The constants in the equation shown in Fig. 3, which expresses $z$ in terms of $\omega$, are given by:

$$a = (n + (n^2 - 1)^{\frac{1}{2}})/\pi,$$
$$b = (n - (n^2 - 1)^{\frac{1}{2}})/\pi,$$
$$c = -(((a+b)/2)\log(a-b) - (ab)^{\frac{1}{2}}\log(a-b)).$$

The transformation given in Fig. 3 allows the flux on the median plane to be calculated for various values of $n$. Changing the value of $n$

---

[6] See for example: W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company Inc., New York, 1939), Chapter IV.

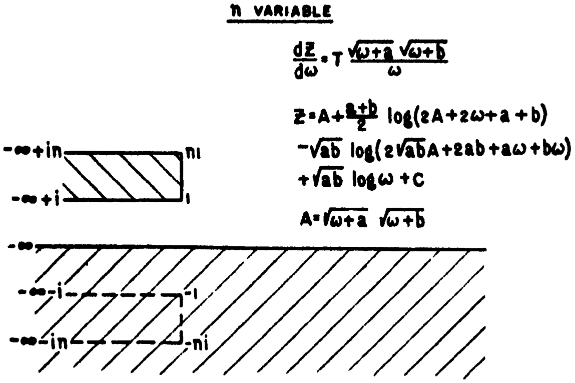

n VARIABLE
$\frac{dz}{d\omega}=T\frac{\sqrt{\omega+a}\sqrt{\omega+b}}{\omega}$
$z=A+\frac{a+b}{2}\log(2A+2\omega+a+b)$
$-\sqrt{ab}\log(2\sqrt{ab}A+2ab+a\omega+b\omega)$
$+\sqrt{ab}\log\omega+C$
$A=\sqrt{\omega+a}\sqrt{\omega+b}$
$-\infty+in$
$ni$
$-\infty+i$
$i$
$-\infty$
$-\infty-i$
$-i$
$-\infty-in$
$-ni$

FIG. 3. Type of magnet pole construction considered and conformal transformation used.

corresponds to changing the thickness of the pole pieces relative to the gap width. The function $h(x)$ has been evaluated for several values of $n$, and the results are shown in Fig. 6.

In addition to the evaluation of $h(x)$ for several values of $n$ the two extreme cases of $n=1$ and $n=\infty$ have been treated. Figure 4 shows the boundaries to be transformed and the differential equation for the extreme case of $n=1$. This case which is physically impossible corresponds to the pole pieces being thin sheets of magnetic material. The other extreme case of $n=\infty$, illustrated in Fig. 5, is also physically impossible but may be representative of magnets which have a large ratio of thickness to gap width. The results of the evaluation of $h(x)$ for these two extreme cases are also shown in Fig. 6.

It is somewhat surprising in inspecting Fig. 6 to see how little effect the variation of pole piece thickness has on the fringing flux. In this figure $h(x)$ is plotted against distance from pole-face boundary with the unit of distance as one gap width. Values of $d$ obtained from using these evaluations of $h(x)$ cannot be expected to give true values of $d$ for actual magnets. This is because an actual magnet will have finite dimensions in all directions and for it $h(x)$ will fall to zero more rapidly for large values of distance from pole-face boundary than shown in Fig. 6.

TABLE I. Variation of $d$ with various values of $n$ and $r$.

| $r$ $n$ | 1 | 2 | 5 | 10 | $\infty$ |
|---|---|---|---|---|---|
| 5 | 0.69 | 0.84 | 0.92 | 0.95 | 0.98 |
| 10 | 0.81 | 0.97 | 1.07 | 1.14 | 1.20 |

If the dimensions are, however, reasonably large compared to the gap width, then the obtained values of $h(x)$ may be expected to be fairly accurate throughout the region wherein its greatest reduction in value occurs, i.e., up to about 2 gap widths.

In order to appraise the value of $d$ obtainable for the two-dimensional cases discussed above the necessary integrations as in Eq. (4) were carried out. The results are given in Table I. Here $d$ is given in gap widths as a function of the two variables $n$ and $r$. In each case it applies to the central ray. Actually $d$ is not directly dependent on $r$ but rather on the distance of the ion source from the uniform field region. For the cases evaluated in Table I this distance was taken as equal to $r$, a situation which exists in a sector-type spectrometer when the sector angle is 90°. The values considered were: $r=5$ gap widths and $r=10$ gap widths.

As is evident from Table I, a large change in ion source distance will cause considerable change in $d$. Although the magnetic field strength is small at large distances it has a large cumulative effect. In determining $h(x)$ experimentally for a determination of $d$ by numerical integration it is important to take the measurements under as near operating conditions as possible. For example, if an auxiliary magnet is used with the ion source to align the beam of ionizing electrons the experimental determination of $h(x)$ must include its effects.

It has sometimes been assumed that effective corrections for the influence of the fringing flux could be made by the simple use of the virtual field boundaries. Thus in locating the ion source and ion collector relative to the magnetic field the apex formed by the virtual field boundary is

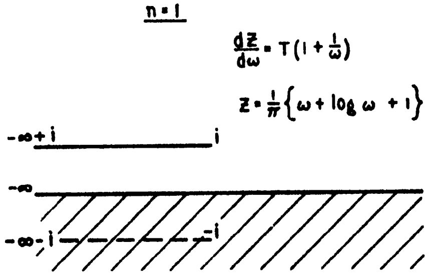

n = 1
$\frac{dz}{d\omega}=T(1+\frac{1}{\omega})$
$z=\frac{1}{\pi}\{\omega+\log\omega+1\}$
$-\infty+i$
$i$
$-\infty$
$-\infty-i$
$-i$

FIG 4. Extreme case of magnet pole construction wherein $n=1$.

obtained by some approximate or graphical method. Such a method of correction assumes that the central ray will have its center of curvature located on the virtual sector apex.

We now wish to show that this method is not exact and that not only must the displacement of virtual field boundaries be taken into account but also the deviation of the trajectory from a straight line in the region between the ion source and the magnetic analyzer.

Figure 7 represents the geometry such as found in a 90° sector-type spectrometer with symmetrical location of ion source and collector. Here $A$ represents the position of an ion source with its axis coincident with line $AB$. The effect of the fringing field is to cause the ions emerging from the source on the line $AB$ to move away from it as they approach the magnetic analyzer. In the scale of dimensions used in constructing this figure $AB$ is 5 gap widths which would be proper for a 90° instrument utilizing trajectories with radii of curvature also of 5 gap widths. The center of curvature for that portion of the central ray within the uniform field region is at $D$, the intersection of the virtual field boundaries $DH$ and $DI$. Now if one were to use $DH$ and $DI$ as the new boundaries of an ideal magnetic field region, the ion source would be located at $F$ which will be 5 gap widths from $DH$ and which lies on the line perpendicular to $ED$, the bisector of the sector angle. However, it is clear that the location of the ion source at $F$ will give a central ray with a center of curvature below and slightly to the right of $D$ as the fringing field will still be operative. Thus the assumption that the use of virtual field boundaries would allow the use of the exact geometrical relations prescribed for this type of instrument (reference 1) is seen not to be true. It is to be noticed that this result does not depend on the numerical value of $d$ or of the sector angle, i.e., the use of virtual field boundaries and the exact geometrical relationships described above will always yield a center of curvature of the central ray offset from the virtual field boundary apex.

A method of correcting for fringing flux which depends upon an assumption concerning optimum focusing is now suggested. It is assumed that optimum focusing conditions are achieved when the ion source and ion collector are located the same distance from the actual field boundaries as would be used for the ideal case wherein there were no fringing flux and are so located relative to the sector apex that the central ray leaves the ion source in a direction normal to one pole-face boundary and enters the ion collector in a direction normal to the other pole-face boundary. Such a condition is illustrated in Fig. 7. There the distances $AB$ and $CG$ are equal to the radius of curvature for the uniform field region which is a condition necessary for the ideal 90° sector-type analyzer. One basis for the above

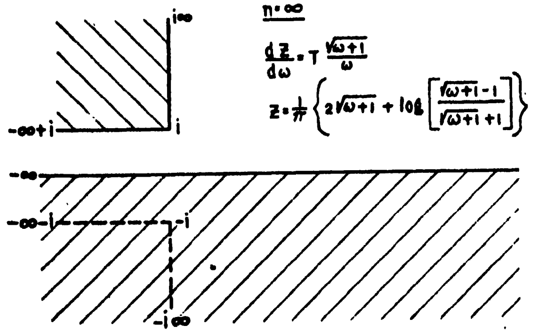

i∞
n=∞
$\frac{dz}{d\omega}=T\frac{\sqrt{\omega+1}}{\omega}$
$z=\frac{1}{\pi}\left\{2\sqrt{\omega+1}+\log\left[\frac{\sqrt{\omega+1}-1}{\sqrt{\omega+1}+1}\right]\right\}$
-∞+i
i
-∞
-∞-i
-i
-i∞

FIG. 5. Extreme case of magnet pole construction wherein $n=\infty$.

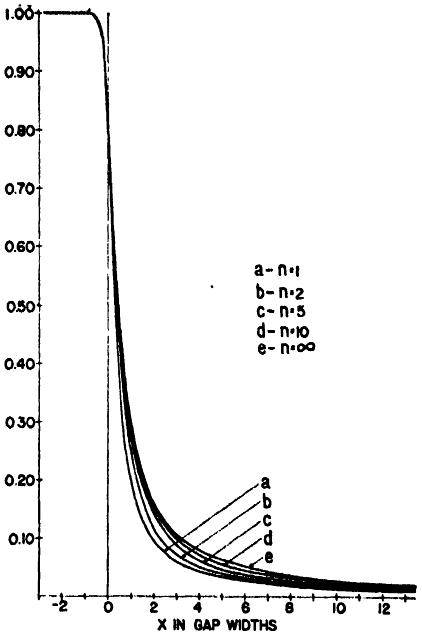


FIG. 6. Fringing magnetic field plotted as a function of distance from magnet pole faces for various values of $n$.

assumption about optimum focusing conditions is the close correspondence between the geometry stated and the geometry of the ideal case which was used in the derivation of the focusing properties of this type of analyzer. Another basis for the assumption is that as an ion beam of small angular spread reaches the actual pole-face boundary, after passing from the ion source through the fringing flux, the angular and spatial separations of the individual trajectories are approximately the same as would be obtained in the ideal case wherein there were no fringing field. Also the centers of curvature of the trajectories as they pass through the uniform field region will be separated by approximately the same distances as for the ideal case. These latter statements will be seen to be true later when we consider the changes in Eqs. (2) and (3) for ion trajectories of different angles of emergence from the source.

By way of demonstrating the above assumption concerning optimum fringing flux corrections let us suppose it is desired to make the corrections for a 90° sector-type instrument with symmetrical location of ion source and collector. Referring to Fig. 7, $AB$ will be the distance of the ion source from the actual field boundary and will be equal to the radius of curvature chosen for operation. Supposing $h(x)$ to be available it will be possible to obtain $f(x)$ and to use it in the integration of Eq. (2), which will give the deviation of the central ray trajectory from the line $AB$ and will trace its entry into the uniform field region. When the trajectory reaches a point deep enough in the uniform field region that $h(x)=1$, the slope $dy/dx$ may be evaluated and the location of point $D$, the center of curvature, determined. Now it is clear that we may expect the ion to leave the field in a trajectory exactly similar to the entry one if symmetrical conditions are provided. To do so a line which is at 45°, half the sector angle, with $DH$, the virtual field boundary, is projected to the actual field boundary. This locates $E$, the actual sector apex. Now it is seen that exactly symmetrical conditions are provided. The dimension of significance now is $EB$ which locates the distances of the ion source and ion-collector axes from the actual sector apex. This is readily obtained in simple steps, using the location of point $D$ relative to points $A$ and $E$ and to the point of entry of line $AB$ into the uniform field region.

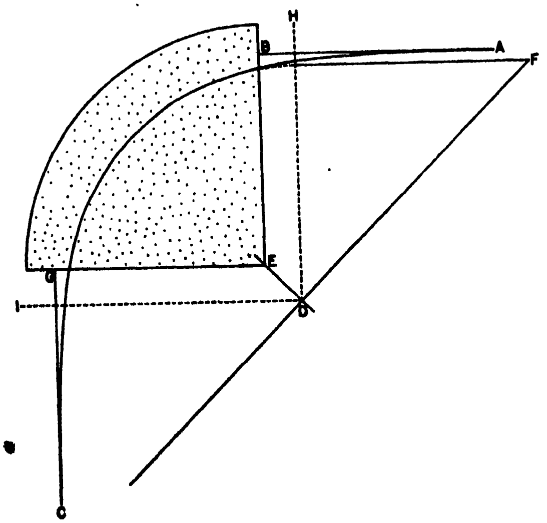


FIG. 7. Geometry of fringing flux corrections for sector-type instrument.

If it is desired to determine the flux corrections for an unsymmetrical system such that the angles $\theta$ and $\gamma$, as defined in reference 1, are unequal it may be done using the methods of above. In this case it would be necessary to determine the trajectories for two fringing field conditions, i.e., for the ion source and ion-collector sides, and match them in the uniform field region.

For the trajectories and geometry considered so far

$$f(x)=\int_{x_0}^{x} h(x)dx,$$

that is $f(x_0)=0$ so that $(dy/dx)x_0=0$. This is the condition imposed on the central ray, i.e., that it leave the ion source normal to the slits (using such a case as in Fig. 7 where the ion source axis is parallel to the $x$ axis and the pole-face bounary is parallel to the $y$ axis). However, for ions emerging from the ion source at a small angle $\alpha$ with the normal to the slits we will have:

$$f(x)=\int_{x_0}^{x} h(x)dx+\tilde{c},$$

where $\tilde{c}$ is a constant to be correlated with the initial slope.[2]

From Eq. (1) we have then:

$$(dy/dx)x_0 = \tan\alpha = \bar{c}/(r^2 - \bar{c}^2)^{\frac{1}{2}}.$$

Since for small $\alpha$ (as would be true for an actual ion beam) $r \gg \bar{c}$ we obtain $\bar{c} = r\tan\alpha$.

The virtual boundary displacement for such a trajectory will be obtained from:

$$r = \pm\int_{x_0}^{\bar{x}} h(x)dx + r\tan\alpha.$$

Since $h(x)$ is equal to unity in the uniform field region we have for the virtual boundary displacement $d(\alpha)$ for a trajectory of starting angle $\alpha$

$$d(\alpha) = d(0) + r\tan\alpha.$$

The separation then between the virtual boundary effective for a trajectory of starting angle $\alpha$ and the virtual boundary effective for the principal ray is seen to be $r\tan\alpha$ which is the same as is obtained for the ideal case. This result is part of the basis for the assumption above concerning optimum focusing conditions.

The separation in the $y$ direction at the pole-face boundary between the central ray and one emerging from the ion source with angle $\alpha$ may be seen to be approximately the same as for the ideal case. To calculate this difference we integrate Eq. (2) for the two cases. We get then for the separation

$$\Delta y = \bar{c}\int_{x_0}^{0} \frac{dx}{(r^2 - (\bar{f} + \bar{c})^2)^{\frac{1}{2}}}.$$

However, since $r \gg \bar{c}$ and since $\bar{f}(x)$ is always found to be small enough so that $(r^2 - (\bar{f} + \bar{c})^2)^{\frac{1}{2}} = r$ to within about 2 percent or better for all of $0 < x < x_0$ we have $\Delta y = \bar{c}x_0/r$ which is equal to $x_0\tan\alpha$, the same value as for the ideal case.

By use of Eq. (2) and the approximations above we may show that the angular spread of the trajectories as they enter the uniform magnetic field region is approximately the same for the actual case as for the ideal one. We shall denote the slope angle at $x = 0$ as $\beta$. Then from Eq. (2) we have:

$$\tan\beta = (\bar{f}(0) + \bar{c})/(r^2 - (\bar{f}(0) + \bar{c})^2)^{\frac{1}{2}}. \qquad (5)$$

By differentiation we get

$$(1 + \tan^2\beta)\Delta\beta = r^2\Delta\bar{c}/(r^2 - (\bar{f}(0) + \bar{c})^2)^{\frac{1}{2}}.$$

Using Eq. (5) and the approximation of above this reduces to $\Delta\beta = \Delta\bar{c}/r$. For the central ray and one emerging from the source at a small angle $\alpha$ we have $\Delta\bar{c} = \bar{c} = r\tan\alpha = r\alpha$ so that $\Delta\beta = \alpha$; i.e., the trajectories will have approximately the same angular divergence at $x = 0$ as at the ion source. An actual calculation of the angular spread at $x = 0$ for one of the cases for which $d$ was evaluated above is illuminating. The conditions used were: $n = 10$, $r = 5$, and $\alpha = 3°40'$. For this case $\bar{f}(x)$ was calculated for the region between the ion source and the actual field boundary using the values of $h(x)$ shown in Fig. 6. From this and from Eq. (5) it was determined that the angle at $x = 0$ between the tangents of the two trajectories, was $3°44'$, i.e., the same as the starting angle to within less than 2 percent discrepancy.

The author is indebted to Dr. M. Muskat for several discussions of this problem and to Dr. P. D. Foote, Executive Vice-President of Gulf Research and Development Company, for permission to publish this material.

# Quarter Wave Coupled Wave-Guide Filters*

Wilbur L. Pritchard

(Received March 24, 1947)

This paper considers the theory of band-pass wave-guide filters formed by cascading rectangular cavities separated by quarter guide wave-lengths. Formulas for resonant frequency, loaded $Q$, and insertion loss in the pass band are derived for a single stage using matrix theory and are in a form suitable for design purposes.

The general theory of "$n$" stages is worked out using matrix algebra, and it is shown that the insertion loss and phase shift of "$n$" stages are given by:

$$L_n = 1 + x^2 Un^2(x) \qquad \phi = \tan^{-1}[Tn(x)/Un(x)],$$

where $Tn(x)$ and $Un(x)$ are the first and rationalized second kind Tschebyscheff polynomials.

The specific design procedure is outlined including the design of a finite thickness inductive aperture.

## I. INTRODUCTION

THE filters considered here are of the band-pass type, formed by rectangular cavities in cascade and separated from each other by quarter wave sections of line. The complete problem is divided into three sections. First, analysis of a single stage to find resonant frequency, phase shift, loaded $Q$, and loss in the pass band; second, study of the behavior of cascaded stages to find insertion loss out of the pass band, phase shift, and loss in the pass band for "$n$" stages; and third, the final design procedure.

The theory of a single stage is handled by use of matrix theory of four terminal networks[1] and the combined lumped element-transmission line equivalent circuit for the cavity. The cavity proper is represented by a length of ideal line, and the coupling apertures are represented by pure shunt susceptances The latter representation is not strictly accurate, and a slight correction is necessary in the final design. The analysis of "$n$" stages in cascade is also handled by matrix theory and involves taking the "$n$th" power of a matrix.

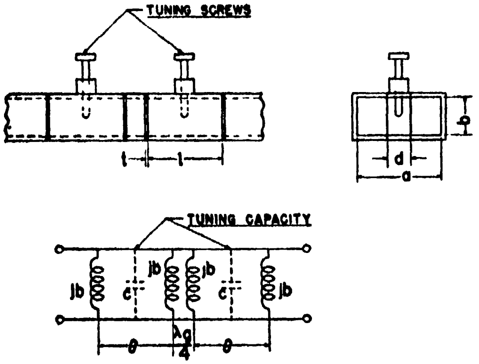


Fig. 1. Two stage filter and network equivalent.

Figure 1 shows the wave-guide configuration along with the equivalent circuit used in the analysis.

## II. PRELIMINARY NETWORK THEORY

A few brief relations based on the general circuit matrix of a quadripole are now derived. If the box in Fig. 2 represents any linear, passive network we can write:

$$E_1 = AE_2 + BI_2, \tag{1}$$

$$I_1 = CE_2 + DI_2. \tag{2}$$

The coefficients $A$, $B$, $C$, and $D$ completely determine the network behavior. Evidently $A$ and $D$ are numerics, $B$ is an impedance, and $C$ is an admittance. Actually there are only three independent coefficients, since application of the reciprocity theorem to Eqs. (1) and (2) yields the relation.

$$AD - BC = 1. \tag{3}$$

Furthermore in the case of symmetrical networks we also have

$$A = D. \tag{4}$$

---

* The author wishes to acknowledge his indebtedness to R. M. Fano of M.I.T. for a preview of the manuscript of Vol. XXII of the Rad. Lab. Series which deals with the same problem from a different approach.

[1] L. A. Pipes, "Matrix theory of four terminal networks," Phil. Mag. 30, 370–395 (November 1940); L. Brillouin, Revue Generale de L'Electricite 55, 3–16 (January 4, 1946).

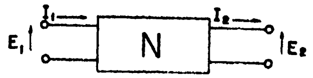


FIG. 2.

Since all the networks considered in the subsequent analysis are symmetrical, Eq. (4) will prove quite useful.

Rewriting Eqs. (1) and (2) in matrix form yields:

$$\begin{bmatrix} E_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} E_2 \\ I_2 \end{bmatrix}. \quad (5)$$

The matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is called the general circuit matrix and will henceforth be designated as $[u]$. The general circuit matrices for several basic network elements used in this paper are shown in Fig. 3.

In Fig. 4 we have a generator and load, normally matched to each other, separated by some four terminal network. The normalized generator and load impedances are assumed to be real and equal to one, since this is the condition commonly met in wave-guide problems. Since the impedances are all normalized, the following equations will not be dimensionally homogeneous.

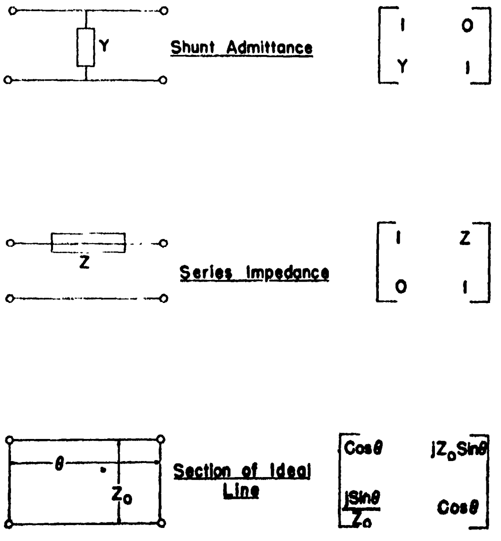


FIG. 3. General circuit matrices of several common networks.

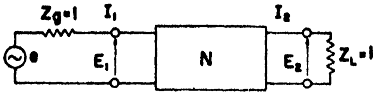


FIG. 4.

From Eqs. (1) and (2) we have:

$$Z_1 = E_1/I_1 = (AZ_L+B)/(CZ_L+D) = (A+B)/(C+D). \quad (6)$$

Using simple network theory we can find $R$, the ratio of the voltages across $Z_L$ when directly connected to the generator and when connected through the network,

$$R = \tfrac{1}{2}(A+B+C+D). \quad (7)$$

$R$ will be a complex quantity in general. The insertion loss, defined as the ratio of available power to the power delivered through the network, is found from the magnitude of $R$.

$$L = |R|^2. \quad (8)$$

The insertion phase shift, or phase shift in the voltage across the load resulting from the insertion of the network, is found from the real and imaginary components of $R$.

$$\tan\phi = [I(R)/R_e(R)]. \quad (9)$$

Equations (7)–(9) will enable us to determine the attenuation and phase characteristics of both single and multiple stage filters from the general circuit matrix.

## III. ANALYSIS OF A SINGLE STAGE

Figure 5 shows the equivalent circuit for a single stage. We proceed with the analysis of a single stage by first determining the general circuit matrix. Since the circuit of Fig. 5 can be considered as a cascade combination of two shunt susceptances and a section of ideal line we find $[u]$ by matrix multiplication in the same order as the order of connection.

Referring to the table in Fig. 3

$$[u] = \begin{bmatrix} 1 & 0 \\ jb & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & j\sin\theta \\ j\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ jb & 1 \end{bmatrix}. \quad (10)$$

Performing the indicated multiplication we have:

$$[u] = \begin{bmatrix} \cos\theta - b\sin\theta & j\sin\theta \\ j[(1-b^2)\sin\theta + 2b\cos\theta & \cos\theta - b\sin\theta \end{bmatrix}. \quad (11)$$

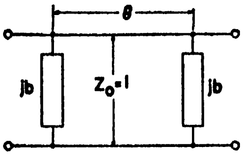


FIG. 5. Network equivalent of single stage filter.

From Eq. (7) we find $R$

$$R=(\cos\theta-b\sin\theta)$$
$$+j[(1-b^2/2)\sin\theta+b\cos\theta]. \quad (12)$$

Taking the sum of the squares of the components of $R$ and simplifying gives for the insertion loss:

$$L=1+b^2[\cos\theta-(b/2)\sin\theta]^2 \quad (13)$$

and for phase shift:

$$\tan\phi=\frac{[1-(b^2/2)]\sin\theta+b\cos\theta}{\cos\theta-b\sin\theta}. \quad (14)$$

Since the network is lossless a first definition for the resonant frequency is that frequency at which $L=1$. We have from Eq. (13)

$$1=1+b^2[\cos\theta-(b/2)\sin\theta]^2$$

Solving for $\theta$

$$\theta=\text{arc}\tan(2/b). \quad (15)$$

Equation (15) determines the resonant length of the filter in terms of the aperture susceptance. For large values of $b$ which represent physically small apertures, $\theta$ is almost 180° or one-half wave-length. Equation (15) is plotted in Fig. 6, curve (2).

A second definition of resonance, and one which will be useful in the subsequent determination of the filter $Q$, is that frequency at which the input admittance is zero if the load admittance is zero. From Eq. (6) we have for the input admittance when $Y_L=0$

$$Y_1=C/A.$$

Substituting from the matrix in Eq. (11) yields

$$Y_1=j\frac{(1-b^2)\sin\theta+2b\cos\theta}{\cos\theta-b\sin\theta}.$$

Equating the numerator to zero we have:

$$(1-b^2)\sin\theta=-2b\cos\theta$$
$$\tan\theta=-2b/(1-b^2)$$
$$\theta=\text{arc}\tan[2b/(b^2-1)]. \quad (16)$$

Equations (15) and (16) provide two different design relations for resonant frequency. For values of $b$ corresponding to values of $Q$ usually met in practice (treated in next section) the two formulas give the same result for $\tan\theta$ within normal slide-rule discrepancies.

To determine the physical length $l$ of the cavity we use the relation

$$\theta=\beta l, \quad \text{where} \quad \beta=2\pi/\lambda_g. \quad (17)$$

$\beta$ is the propagation constant of the guide and $\lambda_g$ is the guide wave-length.

Thus

$$\tan 2\pi l/\lambda_g=2/b \text{ where } \lambda_g=\lambda/[1-(\lambda/\lambda_c)^2]^{\frac{1}{2}}. \quad (18)$$

In terms of $\omega$, the angular velocity we have:

$$\tan\frac{\omega_0 l}{c}\left[1-\left(\frac{\omega_c}{\omega_0}\right)^2\right]^{\frac{1}{2}}=2/b, \quad (19)$$

where $l$ = cavity length, $b$ = aperture susceptance, $\lambda$ = air wave-length, $\lambda_c$ = cut-off wave-length, $\omega_0$ = resonant angular velocity, $\omega$ = cut-off angular velocity, and $c$ = velocity of light in vacuum.

Equation (19) is a design equation for determining the length of the resonant cavity. However, since there is an infinite number of pairs of values of $b$ and $l$ which satisfy Eq. (19) a second condition is required. This is the relation between the operating or loaded $Q$ of the cavity and the aperture susceptance. To derive such a relation we find a lumped parallel resonant cir-

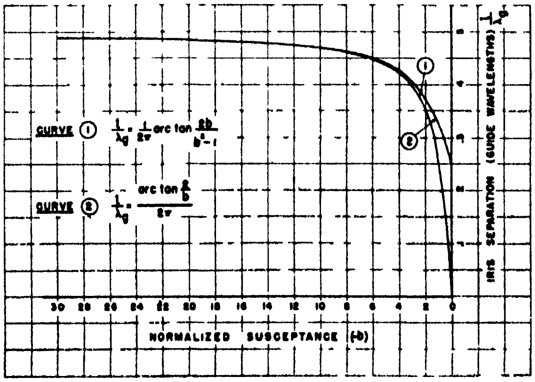


FIG. 6.

cuit which closely approximates the filter's behavior in the vicinity of resonance. In determining the filter's behavior near resonance we assume that the frequency consciousness of the apertures is negligible compared to the frequency consciousness of the length of line. In other words we assume constant $b$ and consider variations in $\theta$ with frequency and their effects on the filter characteristics. Physically this assumption corresponds to neglecting the effect of the coupling holes in considering the characteristics *versus* frequency of a resonant cavity. Naturally this assumption is most valid for small coupling holes or large susceptances. Subsequently it will be seen that for usual values of $Q$ the aperture susceptance will be high enough to make the above assumption quite justifiable.

Referring to Fig. 3 and the general circuit matrix we have for the input impedance, when the load impedance is infinite, from Eq. (1) and (2)

$$Z_1=(A+B/Z_L)/(C+D/Z_2)=A/C.$$

On an admittance basis we have from Eq. (11)

$$Y_1=C/A=j\frac{2b+(1-b^2)\tan\theta}{1-b\tan\theta}. \qquad (20)$$

We assume $\theta=\theta_0+K\Delta f$, where $\theta_0$ is the resonant wave-length and $K=d\theta/df$. By using the identity for the tangent of the sum of two angles and making the approximations $\tan K\Delta f=K\Delta f$ and $bK\Delta f<1$ Eq. (20) can be reduced to the form:

$$Y_1=jK\Delta f(b^2+1). \qquad (21)$$

In Fig. 7 the input admittance with zero load admittance is the well-known relation:

$$Y_1=j(C/L)^{\frac{1}{2}}(\omega/\omega_0-\omega_0/\omega), \qquad (22)$$

or for small frequency deviations $(\omega/\omega_0-\omega_0/\omega)=2\Delta\omega/\omega_0=\epsilon$,

$$Y_1=j(C/L)^{\frac{1}{2}}\epsilon. \qquad (23)$$

Equating the two input admittances we have:

$$j(C/L)^{\frac{1}{2}}2\Delta f/f_0=jK\Delta f(b^2+1) \qquad (24)$$

$$(C/L)^{\frac{1}{2}}=(Kf)(b^2+1)/2. \qquad (25)$$

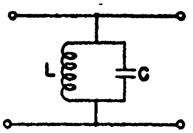


FIG. 7.

If we take the cavities as non-dissipative, then the $Q$ for any given type of aperture depends only on the loading. Since there is a matched load $Z_0=1$ on each side of the cavity the shunt loading resistance in the equivalent resonant circuit is taken equal to $\frac{1}{2}$. Since

$$\begin{aligned}R_{\text{shunt}}&=Q(L/C)^{\frac{1}{2}}=\tfrac{1}{2}\\ Q&=\tfrac{1}{2}(C/L)^{\frac{1}{2}} \qquad (26)\\ Q&=(Kf)(b^2+1)/4.\end{aligned}$$

To find $K$ we take the first derivative of the function relating $\theta$ to $f$

$$\theta=\frac{2\pi lf}{c}[1-(fc/f)^2]^{\frac{1}{2}}. \qquad (27)$$

After differentiating and simplifying

$$K=\frac{d\theta}{df}=\frac{2\pi l}{c}[1-(fc/f)^2]^{-\frac{1}{2}}. \qquad (28)$$

Also, from Eq. (27)

$$f_0=\frac{c\theta_0}{2\pi l}[1-(fc/f)^2]^{-\frac{1}{2}}. \qquad (29)$$

Substituting Eqs. (28) and (29) in Eq. (26) we find:

$$Q=(b^2+1)\theta_0/4[1-(fc/f)^2], \qquad (30)$$

but

$$\theta_0=\tan^{-1}2b/(b^2-1).$$

Therefore:

$$Q=\frac{b^2+1}{4[1-(fc/f)^2]}\tan^{-1}\frac{2b}{b^2-1}. \qquad (31)$$

For large values of $b$

$$\tan^{-1}\frac{2b}{b^2-1}=\pi \quad\text{and}\quad b^2\gg 1.$$

$$Q=\frac{b^2}{4[1-(fc/f)^2]}. \qquad (32)$$

Equation (31) is plotted in Fig. 8. The calculation of $Q$ using this expression neglects the effects of dissipative losses in the cavity. These

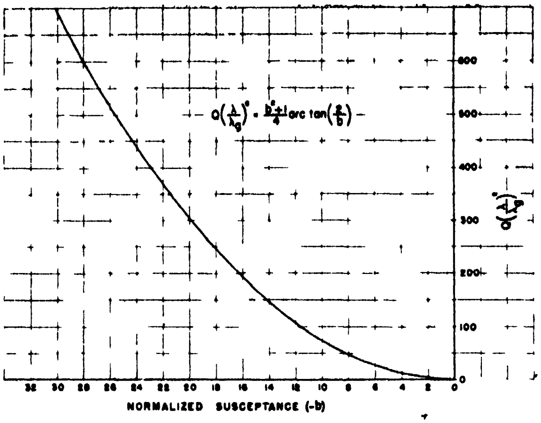


FIG. 8.

losses will have two chief effects on the filter performance. One will be to produce a slight lowering of the loaded $Q$ according to

$$1/Q^1 = 1/Q + 1/Q_0, \qquad (33)$$

where $Q^1$ is the actual operating $Q$, and $Q_0$ is the unloaded $Q$ of the rectangular cavity which can be calculated from ordinary resonant cavity theory.[2] The second effect will be to produce losses in the pass band. These losses can also be calculated from the unloaded $Q$ in the following manner:

From the general circuit matrix for a shunt admittance we have:

$$R = (A+B+C+D)/2 = (2+Y)/2$$
$$R = 1 + Y/2. \qquad (34)$$

For a dissipative cavity $Y$ is equal to the equivalent shunt conductance at resonance or:

$$Y = G_0 = (1/Q_0)(L/C)^{\frac{1}{2}}. \qquad (35)$$

But $(L/C)^{\frac{1}{2}} = \frac{1}{2}Q$, where $Q$ is the loaded $Q$ (Eq. (26)). Therefore

$$G_0 = 2Q/Q_0$$

and

$$R = 1 + Q/Q_0$$
$$L = |R|^2 = (1+Q/Q_0)^2 \qquad (36)$$

or

$$L_{\text{db}} = 20 \log(1+Q/Q_0). \qquad (37)$$

Equations (33) and (37) can be used to correct for the effects of incidental dissipation in the cavity.

## IV. QUARTER WAVE CASCADING OF FILTER SECTIONS

The simple filter of the previous section is only applicable where the desired response is approximately the same as that of a parallel resonant circuit. In many applications this is not sufficient, since the insertion ratio off resonance may not be high enough, or the width of the pass band may not be sufficiently great. In such cases it becomes necessary to cascade several filter sections. This cascading can be done in several ways. Either direct connection can be utilized, or the separate sections can be isolated by quarter wave sections. The successive sections can be tuned to the same frequency, or they can be stagger tuned. This report will consider only quarter wave coupled identical sections.[3]

We continue to use the lumped circuit equivalent for a single stage and consider the case of three quarter wave coupled stages (see Fig. 9).

The associated matrix is given by (refer to chart in Fig. 3):

$$[u] = \begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix} \begin{bmatrix} 0 & j \\ j & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix} \begin{bmatrix} 0 & j \\ j & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix}, \qquad (38)$$

Performing the indicated multiplication we have for $[u]$

$$[u] = \begin{bmatrix} 1+Y^2 & Y \\ Y(2+Y^2) & 1+Y^2 \end{bmatrix}. \qquad (39)$$

In Fig. 10a we have a simple $\pi$ section. To find its associated matrix we again multiply the matrices of its components.

$$[u]_\pi = \begin{bmatrix} (Z_p+Z_s)/Z_p & Z_s \\ (2Z_p+Z_s)/Z_p^2 & (Z_p+Z_s)/Z_p \end{bmatrix}. \qquad (40)$$

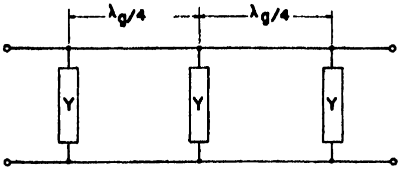


FIG. 9. Network equivalent of three cascaded stages.

---

[2] See, for instance, Sarbacher and Edson, *Hyper and Ultra High Frequency Engineering* (John Wiley and Sons, Inc., New York, 1943).

[3] Quarter wave coupled filters were first developed by R. M. Fano at Rad. Lab.

To find the equivalent $\pi$ section for Fig. 9 we equate the elements of the two matrices given in Eqs. (39) and (40)

$$\begin{aligned} (Z_p+Z_s)/Z_p&=1+Y^2 & Z_s&=Y\\ Z_s/Z_p&=Y^2 & Y_s&=1/Y\\ Z_p=Z_s/Y^2&=1/Y\\ Y_p&=Y. \end{aligned}$$

Using the above relations we have a $\pi$ section lumped circuit equivalent to the three section filter of Fig. 9. The value of $Y$ is given by Eq. (23).

$$\begin{aligned} Y&=j(C/L)^{\frac{1}{2}}\epsilon=Y_p\\ Y_s&=1/Y=-j(L/C)^{\frac{1}{2}}/\epsilon. \end{aligned} \tag{41}$$

The admittance of a lossless series resonant circuit is given by:

$$\begin{aligned} Y_{\text{series}}&=1/j(L^1/C^1)^{\frac{1}{2}}\epsilon\\ &=-j(C^1/L^1)^{\frac{1}{2}}/\epsilon. \end{aligned} \tag{42}$$

By comparing Eq. (42) to Eq. (41) we see that $Y_s$ can be replaced by a series resonant circuit where $C^1=L$ and $L^1=C$. Therefore we have for the final lumped equivalent the circuit of Fig. 10b.

This equivalent is evidently a band-pass filter with a center frequency given by $W_0=1/(LC)^{\frac{1}{2}}$ at which both series and shunt arms are resonant.

Similar lumped equivalents can be developed for any number of cascaded sections.

### V. THEORY OF "$n$" SECTIONS IN CASCADE

The brief considerations of Section IV served to yield an equivalent lumped circuit for the particular case of three cascaded stages. The general case, to be considered now, involves "$n$" stages, and we propose to obtain an exact expression for insertion loss and phase shift. We proceed with this problem by considering the typical stage in a quarter wave coupled cascade. Fig. 11 shows the resonant circuit equivalent of the cavity with eighth wave lines on each side. Evidently a cascade of this type section will result in resonant circuits separated by quarter wave lines.

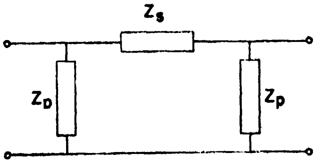


FIG. 10a.

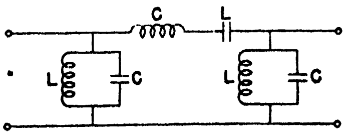


FIG. 10b. Lumped circuit equivalent of three cascaded stages.

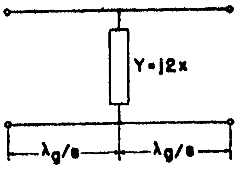


FIG. 11.

Again referring to Fig. 3 we have for the general circuit matrix:

$$[u]=\begin{bmatrix}\cos45° & j\sin45°\\ j\sin45° & \cos45°\end{bmatrix}\begin{bmatrix}1 & 0\\ Y & 1\end{bmatrix}\times\begin{bmatrix}\cos45° & j\sin45°\\ j\sin45° & \cos45°\end{bmatrix}. \tag{43}$$

Substituting $Y=j2X$ where $X=Q\epsilon$, $\cos45°=\sin45°=(2^{\frac{1}{2}}/2)$, and performing the matrix multiplication yields

$$u=\begin{bmatrix}-X & j-jX\\ j+jX & -X\end{bmatrix}. \tag{44}$$

To raise the matrix $[u]$ to the $n$th power we must reduce it to the form

$$[u]=[K]\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}[K]^{-1}. \tag{45}$$

In this form $[u]^n$ becomes:

$$[u]^n=[K]\begin{bmatrix}\lambda_1{}^n & 0\\ 0 & \lambda_2{}^n\end{bmatrix}[K]^{-1}. \tag{46}$$

$\lambda_1$ and $\lambda_2$ turn out to be the eigenvalues of the matrix $[u]$ and are found from the characteristic equation of $[u]$. The characteristic equation is formed by subtracting $[u]$ from the scalar matrix $[\lambda]$ and equating the principal determinant of the difference to zero. The roots of the

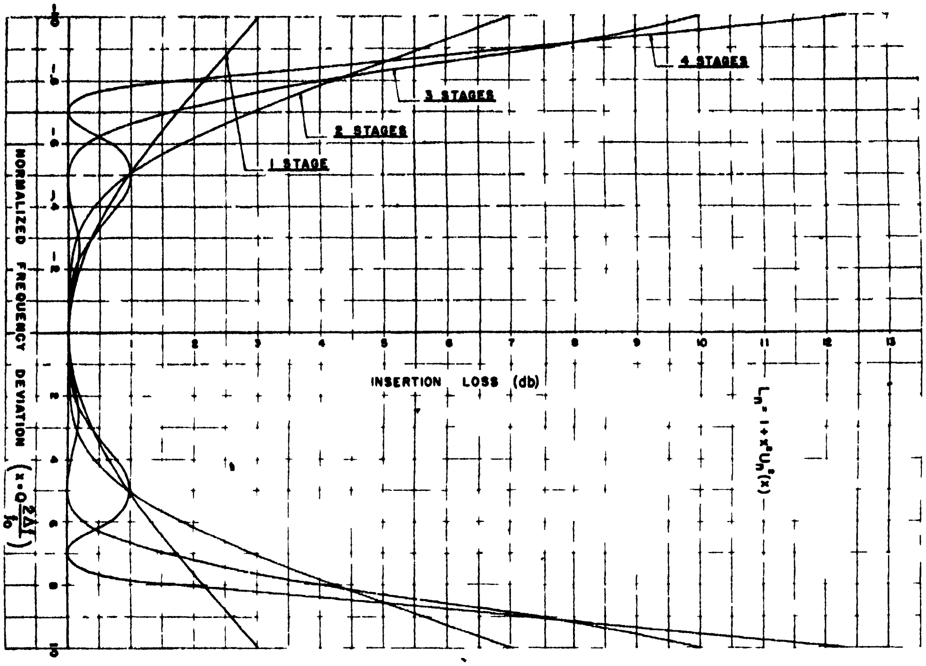


FIG 12

resultant quadratic (Eq. 47) then become:

$$\begin{gathered}\lambda^2+2x\lambda+1=0,\\ \lambda_1=-x+(x^2-1)^{\frac{1}{2}},\quad \lambda_2=-x-(x^2-1)^{\frac{1}{2}},\end{gathered}\tag{47}$$

Also:

$$\lambda_1+\lambda_2=-2x,\quad \lambda_1\lambda_2=1$$

We make the substitutions $\lambda_1=\epsilon^a$ and $\lambda_2=\epsilon^{-a}$ which give:

$$\lambda_1+\lambda_2=2\cosh a,\quad a=\cosh^{-1}(-x).\tag{48}$$

The matrix $[u]$ can now be written in the form:[4]

$$[u]=\begin{bmatrix}\lambda_1+x & \lambda_2+x\\ j(1+x) & j(1+x)\end{bmatrix}\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}$$

$$\times\begin{bmatrix}\dfrac{1}{\lambda_1-\lambda_2} & -\dfrac{\lambda_2+x}{j(1+x)(\lambda_1-\lambda_2)}\\ -\dfrac{1}{\lambda_1-\lambda_2} & \dfrac{\lambda_1+x}{j(1+x)(\lambda_1-\lambda_2)}\end{bmatrix}.\tag{49}$$

The first and third matrices in the above product are inverses of each other, while the center matrix is a scalar matrix. That the product in the right-hand member of Eq (49) is actually equal to $[u]$ can be verified by multiplication.

We now raise $[u]$ to the $n$th power by taking the $n$th power of the elements of the diagonal matrix, multiply the three matrices together and simplify to get:

$$[u]^n=\begin{bmatrix}\cosh an & -j\dfrac{\sinh a}{1+x}\sinh an\\ j\dfrac{1+x}{\sinh a}\sinh an & \cosh an\end{bmatrix}$$

$$=\begin{bmatrix}A_n & B_n\\ C_n & D_n\end{bmatrix}.\tag{50}$$

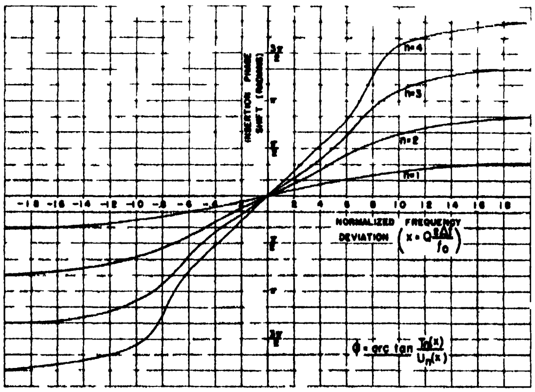


FIG. 13.

---

[4] The transformation and diagonalizing of a matrix carried out above are described in detail in Margenau and Murphy, *Mathematics of Physics and Chemistry* (D. Van Nostrand Company, Inc., New York 1943).

Equation (50) gives the general circuit matrix for "$n$" cascaded stages. Note that $A_n = D_n$ and $A_nD_n - B_nC_n = 1$ which is to be expected. From $[u]^n$ we can now find the general expression for insertion less and phase shift: From Eq. (7) we have:

$$R_n = (A_n + B_n + C_n + D_n)/2.$$

Substituting from the matrix $[u]^n$ yields

$$R_n = \cosh an + j\frac{\sinh an}{\sinh a}. \qquad (51)$$

For insertion loss we take the square of the magnitude:

$$L_n = |R_n|^2 = \cosh^2 an + \sinh^2 an/\sinh^2 a. \qquad (52)$$

Using the identity $\cosh^2 an - \sinh^2 an = 1$ and substituting for Sinh$a$ we can write Eq. (52) in the form:

$$L_n = 1 + x^2\,\mathrm{Sinh}^2 an/(x^2-1) \qquad (53)$$

or

$$L_n = 1 + x^2 \times [\mathrm{Sinh}\{n \operatorname{arc}\cosh(-x)\}]^2/(x^2-1). \qquad (54)$$

The expression in the bracket of Eq. (54) is equivalent to the rationalized Tschebyscheff polynomial of the second kind.[5] Rewriting Eq. (54) we have (see Appendix I):

$$L_n = 1 + x^2 U_n^2(x). \qquad (55)$$

Equation (56) gives the specific polynomials for $U_n(X)$ and $L_n$ for values of "$n$" up to four along with a recurrence formula for $U_n$ in case higher values of "$n$" are desired:

$$\begin{aligned} L_1 &= 1 + x^2, & U_1 &= 1,\\ L_2 &= 1 + 4x^4, & U_2 &= 2x,\\ L_3 &= 1 + x^2 - 8x^4 + 16x^6, & U_3 &= 4x^2 - 1,\\ L_4 &= 1 + 16x^4 - 64x^6 + 64x^8, & U_4 &= 8x^3 - 1, \end{aligned}$$

$$U_{n+1}(x) = 2xU_n(x) - U_{n-1}(x). \qquad (56)$$

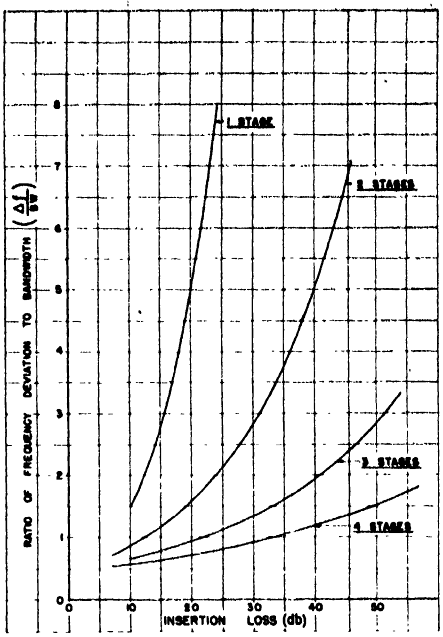


FIG. 14a.

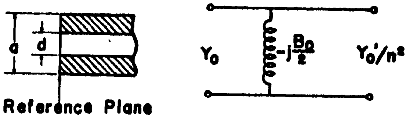


FIG. 14b. Network equivalent of inductive change in wave guide cross section.

Figure 12 is a plot of Eq. (55) in decibels for $n = 1, 2, 3, 4$. From the complex expression for $R_n$ we can also determine the phase shift for "$n$" stages:

$$\phi_n = \tan^{-1}\frac{\sinh an/\sinh a}{\cosh an} = \tan^{-1}\frac{\sinh[n\cosh^{-1}(-x)]/(x^2-1)^{\frac{1}{2}}}{\cosh[n\cosh^{-1}(-x)]}. \qquad (57)$$

The expression in the denominator for Eq. (57) is a Tschebyscheff polynomial of the first kind, while the numerator is again a rationalized second kind polynomial. Rewriting Eq. (57) we have (see Appendix I):

$$\phi_n = \tan^{-1}\frac{-U_n(x)}{T_n(x)}. \qquad (58)$$

Since the typical stage used in considering "$n$" stages cascaded had eight wave line sections on each end, when a particular number of stages is considered there is an extra eighth wave on

---

[5] E. Modelung, *Die Mathematischen Hilfsmittel des Physikers* (Dover Publications, New York, 1943).

each end of the composite filter. This total extra quarter wave introduces an added 90° phase shift. Since this added phase shift is not pertinent to the filter operation we subtract it from the expression in Eq. (58) yielding:

$$\phi_n = \tan^{-1}\frac{-T_n(x)}{U_n(x)}. \tag{59}$$

Specific expressions for $\phi_n$ and $T_n$ for values of "$n$" up to four are given in Eq. (60) along with a recurrence formula for $T_n$

$$\phi_1 = \tan^{-1}x, \qquad T_1 = x,$$

$$\phi_2 = \tan^{-1}(2x^2-1)/2x, \qquad T_2 = 2x^2-1,$$

$$\phi_3 = \tan^{-1}(4x^3-3x)/(4x^2-1), \quad T_3 = 4x^3-3x,$$

$$\phi_4 = \tan^{-1}(8x^4-8x^2+1)/(8x^3-4x),$$

$$T_4 = 8x^4-8x^2+1,$$

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x). \tag{60}$$

Figure 13 is a plot of Eq. (59) for values of "$n$" up to four. In computing phase shift it is assumed that the phase shift at resonance is zero. Actually it will be some multiple of 90° because of the quarter wave sections and the lengths of guide composing the cavities. Also, if the individual stages are designed on an input match basis there will be a very slight extra phase shift at resonance. Generally this is negligible. Since one is usually only interested in phase shift variation about an arbitrary constant value, for f-m distortion considerations, etc., plotting $\phi_n$ about zero is quite useful.

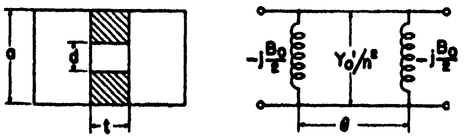


FIG. 15. Network equivalent of finite thickness inductive aperture.

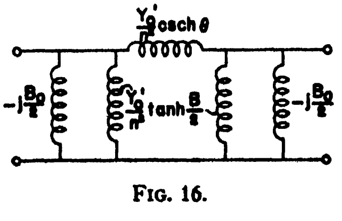

$\frac{Y_0'}{n^2}\operatorname{csch}\theta$
$-j\frac{B_0}{2}$
$\frac{Y_0'}{n^2}\tanh\frac{\theta}{2}$
$-j\frac{B_0}{2}$

FIG. 16.

An approximate relation for the insertion loss of $n$ stages, which will be accurate for values of $x$ far off resonance, can be found from the Tschebyscheff polynomial by neglecting all but the highest power of $x$. This yields:

$$L_{n(\mathrm{db})} = 6(n-1)+20n\log x. \tag{60}$$

Equation (60) can be used to determine if a particular number of stages will have sufficient attenuation far out of the pass band.

## VI. DESIGN PROCEDURE

In designing a filter for a particular application the first step is the choice of the proper number of stages. This is done with the aid of Fig. 14a. The ratio to band width of frequency deviation for a particular attenuation is plotted in Fig. 14 for $n=1, 2, 3, 4$. We define the band width of the filter as the difference between the frequencies of the half-power points. As an example, if an attenuation of 40 decibels is desired at a frequency two band widths removed from resonance, Fig. 14 shows that three stages are necessary. The curves in Fig. 14 can only be used to select the number of stages if band width is considered as distance between the half-power points.

After the number of stages has been determined the $Q$ of a typical stage is found from Fig. 12. That is, from the normalized frequency deviation $X$ which gives the desired band width, and from the band width and resonant frequency, $Q$ is determined using the relation

$$X = Q[(f/f_0)-(f_0/f)] \tag{61}$$

or approximately

$$X = Q(2\Delta f/f_0). \tag{62}$$

The approximation in Eq. (62) is reasonably valid for values of frequency deviation as high as 10 percent which, at microwave frequencies, is much larger than normally considered.

Once $Q$ is chosen and knowing $\lambda/\lambda_g$ the aperture susceptance is found using Fig. 8, and then the aperture separation using Fig. 6. From the guide wave-length the actual mechanical cavity length is readily determined.

The last step in the design procedure is the determination of the actual aperture opening which will produce the desired susceptance.

Unfortunately this is a rather tedious process, since it is absolutely necessary to consider the finite thickness of the window.

The method is based on the equivalent circuit for a symmetrical change in cross section of a rectangular guide. This equivalent was derived by N. Marcuvitz at Radiation Lab.[6] Figure 14b shows the wave guide configuration, and the network equivalent referred to the indicated reference plane:
"$a$" is the cut-off controlling dimension of the guide, $Y_0$ the characteristic admittance of the first section and $Y_0^1/n^2$ the characteristic admittance of the second section of width $d$. $B_0$ is the susceptance of an infinitely thin aperture of opening "$d$" and can be found from Eq. (63) or from curves in the Wave Guide Handbook.

$$B_0/Y_0=\frac{\lambda_g}{a}\left[\cot^2\frac{\pi d}{2a}\right], \qquad (63)$$

$$n^2=\frac{4}{\pi}\cdot\frac{\cos\pi d/2a}{1-(d/a)^2}, \qquad (64)$$

$$\lambda_g'=j2a/(1-(2d/\lambda)^2)^{\frac{1}{2}}, \qquad (65)$$

$$Y_0'=Y_0[-j\lambda_g a/|\lambda_g'|d]. \qquad (66)$$

$|\lambda_g^1|$ is the magnitude of the wave-length in the below cut-off section of guide of width $d$. Equation (65) for the imaginary wave-length in the cut-off guide is the usual expression for guide wave-length modified to take care of $(\lambda/\lambda_c)^2$ greater than one.

The above equivalent circuit can be extended to the case of a finite thickness aperture. Figure 15 shows the wave guide configuration and the equivalent circuit for symmetric inductive apertures.

In the above network the two lumped susceptances represent the abrupt changes in cross section at each end of the window, and the line section represents the short length of below cut-off line produced by the thickness of the window. This line section can be replaced by its exact $\pi$ equivalent,[7] as shown in Fig. 16.

$\theta$ is given for a dissipationless line by Eq. (67)

$$\theta=j\beta l=\frac{2\pi}{|\lambda_g'|}t. \qquad (67)$$

The network of Fig. 16 can be further simplified by combining the shunt susceptances of the $\pi$ equivalent with $-jB_0/2$. The equivalent for the composite aperture is shown in Fig. (17).

$$B_a=B_0/2+|Y_0'/n^2|\tanh\pi t/|\lambda_g'|, \qquad (68)$$

$$B_b=|Y_0'/n^2|\ \operatorname{csch}2\pi t/|\lambda_g'|. \qquad (69)$$

Equations (68) and (69) completely determine the equivalent network for a finite thickness symmetric inductive aperture. To proceed with the design of the filter it is convenient to convert this $\pi$ section to an equivalent $T$. This equivalent is shown in Fig. (18).

$$B_1=2B_b+B_a, \qquad (70)$$

$$B_2=2B_a+B_a^2/B_b. \qquad (71)$$

The shunt arm ($B_2$) of the $T$ equivalent may now be used in determining the $Q$ of the filter stage, and the series arm $B_1$ is used to make a slight correction in the resonant length of the filter. This correction is made by simply shortening the cavity an amount equal to the line length represented by the series susceptance $B_1$.

As a practical consideration it is generally desirable to design the filter at a frequency slightly higher than desired and then tune it down to the desired frequency with a capacitive tuning screw as shown in Fig. 1. This capacity will lower the resonant frequency as much as 5 percent without seriously changing the loaded $Q$.

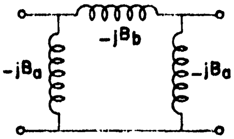


FIG. 17.

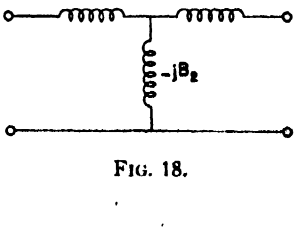


FIG. 18.

---

[6] *Wave Guide Handbook*, Supplement Rad. Lab. Rep. 41, Sec. 21a (January 23, 1945).

[7] Terman, *Radio Engineers Handbook.*

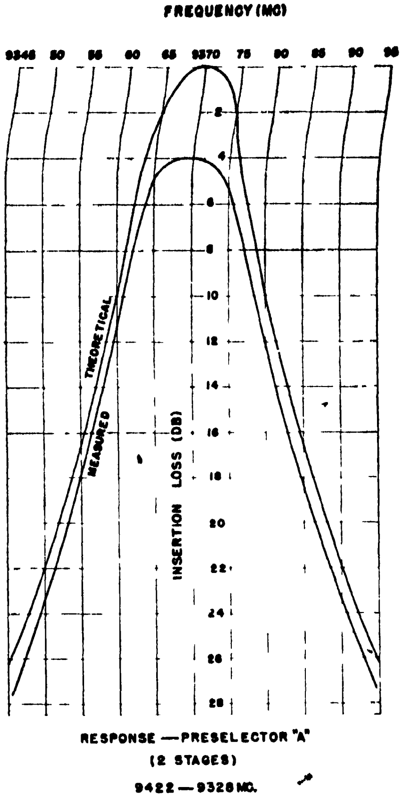


FIG. 19.

## VII. EXPERIMENTAL RESULTS

Figure 19 shows a comparison between the calculated and measured characteristics *vs.* frequency of a two stage filter designed using the method outlined. The filter was designed for a frequency of 9422 Mc in 1″×½″ wave guide and then tuned down to 9370 Mc with two capacitive probes mechanically ganged. The experimental curve is displaced downward from theoretical because of finite losses in the pass band, but its shape is substantially unaltered. It is important to note that the filter can easily be tuned 50 Mc below its natural frequency without materially altering its response.

## Conclusions

A general theory for an "*n*" stage filter has been developed using matrix theory which, along with the design formulas for a single stage and the procedure outlined in the last section, can be used to design a composite filter.

## APPENDIX I

We define the Tschebyscheff function of the first kind as follows:

$$T_n(x)=\cosh(n\cosh^{-1}x), \qquad (1)$$

Letting $y=\cosh^{-1}x$ we have:

$$T_n(x)=\cosh ny \qquad (2)$$

$$T_{n+1}(x)=\cosh(ny+y)=\cosh ny\cosh y+\sinh ny\sinh y \qquad (3)$$

$$T_{n-1}(x)=\cosh(ny-y)=\cosh ny\cosh y-\sinh ny\sinh y. \qquad (4)$$

Adding Eqs. (3) and (4) yields:

$$\begin{aligned}T_{n+1}(x)&=2\cosh ny\cosh y-T_{n-1}(x)\\ T_{n+1}(x)&=2T_n(x)T_1(x)-T_{n-1}(x).\end{aligned} \qquad (5)$$

Equation (5) is a recursion formula from which the Tschebyscheff polynomials can be formed, starting with $T_0(x)=1$ and $T_1(x)=x$. If $-x$ is substituted for $x$ then the sign of $T_n(x)$ changes for $n$ odd and remains the same for $n$ even. In Eq. (57) for phase shift the function in the denominator is actually $T_n(-x)$.

The second kind Tschebyscheff function is defined as follows:

$$U_n(x)=\frac{\sinh(n\cosh^{-1}x)}{(x^2-1)^{\frac{1}{2}}}. \qquad (6)$$

The term $(x^2-1)^{\frac{1}{2}}$ is a rationalizing factor since $U_1(x)=\sinh(\cosh^{-1}x)=(x^2-1)^{\frac{1}{2}}$ and polynomials formed from a recursion formula will always have this radical as a multiplier. The recursion formula for $U_n(x)$, derived in exactly the same manner as the one for $T_n(x)$, is given by:

$$U_{n+1}(x)=2xU_n(x)-U_{n-1}(x). \qquad (7)$$

If $-x$ is substituted for $x$ then the sign of $U_n(x)$ changes for $n$ even and remains the same for $n$ odd. In Eq. (54) giving the insertion loss for an "$n$" stage filter the rationalizing factor is already present in the denominator of the second term of the right-hand member, and since this term is squared the changes in sign of $U_n(-x)$ have no effect and we can write

$$U_n^2(x)=\left[\frac{\sinh(n\cosh^{-1}-x)}{(x^2-1)^{\frac{1}{2}}}\right]^2. \qquad (8)$$

In Eq. (57) for phase shift the inverse tangent of $\phi$ is given by $(U_n(-x)/T_n(-x))$, which is equal to $-(U_n(x)/T_n(x))$ for $n$ either odd or even. Since only relative phase shift is of interest this minus sign has been neglected in plotting $\phi$.

# A Computational Method Applicable to Microwave Networks

R. H. Dicke
*Palmer Physical Laboratory, Princeton University, Princeton, New Jersey*

(Received March 20, 1947)

A method is devised for computing the properties of a complex microwave network in terms of the properties of the circuit elements which in combination form the network. It is particularly suited to machine computation of the properties of circuits of such complexity that simpler, more direct methods fail. It is also applicable to low frequency networks. The elements of such networks may be regarded as interconnected by transmission lines of zero length. A numerical example is used to illustrate the method.

## INTRODUCTION

In the past the engineering of microwave circuits has consisted largely of a method of cut and try. It seems likely that microwave circuits of the future will consist of combinations of microwave elements each of whose properties are known. These microwave elements will probably be interconnected by transmission lines to form complex networks. Among the microwave elements used to synthesize a complex network one may mention tuners, line stretchers, tees, magic tees, directional couplers, and terminations. Figure 1 shows a block diagram of such a microwave circuit.

It is desired to calculate the properties of such a complex network in terms of the known properties of the various sub-elements which in combination form the network. In particular it is desired to calculate the amplitudes and phases of the waves leaving the network at each of the five feeding transmission lines in terms of the amplitudes and phases of the waves entering the network. For some purposes it is also necessary to know the amplitudes and phases of waves in the various interconnecting transmission lines.

As is well known, the complex amplitudes of waves leaving a junction of one or more transmission lines are linearly related to the amplitudes of waves entering the junction. This linear relation can be conveniently expressed by means of a matrix algebra formulation. See Eq. (1). The matrix algebra formulation will be used throughout the remainder of the discussion without going into a discussion of matrix algebra per se. (A concise development of matrix algebra can be found in Margenau and Murphy.[1])

$$\begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} S_{11} & \cdots & S_{1n} \\ \cdot & & \cdot \\ \vdots & & \vdots \\ S_{n1} & \cdots & S_{nn} \end{pmatrix} \begin{pmatrix} a_1 \\ \cdot \\ \vdots \\ a_n \end{pmatrix}. \tag{1}$$

Equation (1) is concerned with a junction of $n$-transmission lines. These lines are numbered consecutively from 1 through $n$. Reference planes are chosen in each of these lines. The complex number $a_j$ represents the amplitude and phase of the wave incident on the junction in the $j$th line, the phase being measured at the reference plane. In a similar way $b_j$ is a complex number representing a wave leaving the junction. The square matrix with elements $S_{jk}$ is called the scattering matrix of the junction. Equation (1) expresses the fact that the $b$'s are linearly related to the $a$'s; Eq. (1) may be written formally as

$$\mathbf{b} = \mathbf{Sa}. \tag{2}$$

It should be noted that $\mathbf{a}$ and $\mathbf{b}$ in Eq. (2) are descriptions of the waves entering and leaving the junction, but that the matrix $\mathbf{S}$ is dependent only on the properties of the junction. The scattering matrix is simply related to the more familiar impedance and admittance matrices.[2]

## DISCUSSION

The problem to be considered is the following: Given a set of microwave circuit elements, with scattering matrices $\mathbf{S}_1, \mathbf{S}_2, \cdots \mathbf{S}_n$, interconnected to form a complex network, what is the scattering matrix of the combined network? A

---

[1] H. Margenau and G. P. Murphy, *The Mathematics of Physics and Chemistry* (D. Van Nostrand Company, Inc., New York, 1943).

[2] Radiation Laboratory Technical Series, *The Principles of Microwave Circuits* (McGraw-Hill Book Company, Inc., New York, to be published), gives a discussion of the scattering matrix formalism and its application to microwave problems.

method will be presented for the calculation of the scattering matrix of a network composed of two interconnected junctions. It is clear that by iteration the method may be used to calculate a network with an arbitrary number of interconnected junctions.

*Problem*: To calculate the scattering matrix of two interconnected microwave elements.

*Element No. 1* has scattering matrix $\mathbf{S}_1$ and is fed by $p$ transmission lines.

*Element No. 2* has scattering matrix $\mathbf{S}_2$ and is fed by $q$ transmission lines.

Let there be $r$ interconnections between the two elements. Number the transmission lines feeding the element (1) in such a way that the terminals of junction (1) number from 1 to $p$. Let the terminals of element (2) be numbered to run from $p-r+1$ to $p-r+q$. Let terminals in the range

$$p-r<n<p+1$$

be interconnected. For an example see Fig. 2. The scattering matrix for element (1) can be written

$$\mathbf{S}_1=\left[\begin{array}{c|c|c} & & 0\\ \hline & & 0\\ \hline 0 & 0 & 0\\ \underbrace{\qquad}_{\substack{p-r\\ \text{columns}}} & \underbrace{\qquad}_{\substack{r\\ \text{columns}}} & \underbrace{\qquad}_{\substack{q-r\\ \text{columns}}}\end{array}\right]\begin{array}{l}\}\,p-r\text{ rows}\\ \}\,r\text{ rows.}\\ \}\,q-r\text{ rows}\\ \phantom{0}\end{array}\qquad(3)$$

Note that zeros have been added to make $\mathbf{S}_1$ a square matrix of order $p-r+q$. Equation (2) applied to junction (1) will be written as

$$\mathbf{S}_1\mathbf{a}_1=\mathbf{b}_1.\qquad(4)$$

In a similar way

$$\mathbf{S}_2\mathbf{a}_2=\mathbf{b}_2,\qquad(5)$$

where

$$\mathbf{S}_2=\left[\begin{array}{c|c|c} 0 & 0 & 0\\ \hline 0 & & \\ \cline{2-2} 0 & & \\ \underbrace{\qquad}_{p-r} & \underbrace{\qquad}_{r} & \underbrace{\qquad}_{q-r}\end{array}\right]\begin{array}{l}\}\,p-r\\ \}\,r\\ \}\,q-r\\ \phantom{0}\end{array}\qquad(6)$$

Introduce an auxiliary matrix

$$\mathbf{A}=\left[\begin{array}{c|c|c} 0 & 0 & 0\\ \hline 0 & \begin{matrix}1_1 & & 0\\ & \ddots & \\ 0 & & 1\end{matrix} & 0\\ \hline 0 & 0 & 0\end{array}\right].\qquad(7)$$

It is characterized by having zeros everywhere but along the diagonal of the central block. Also introduce

$$\mathbf{B}=\mathbf{I}-\mathbf{A}\qquad(8)$$

where $\mathbf{I}$ is the identity matrix.

The matrix $\mathbf{A}$ is useful in representing the connections between the two elements of the network. To see this, note that in Fig. 2 waves leaving junction (1) along transmission lines (4), (5) and (6) are also waves entering junction (2) and *vice versa*. This may be expressed formally in the following way:

$$\mathbf{A}\mathbf{a}_1=\mathbf{A}\mathbf{b}_2,\qquad(9)$$

$$\mathbf{A}\mathbf{b}_1=\mathbf{A}\mathbf{a}_2.\qquad(10)$$

From Eqs. (4), (5), (8)–(10) one obtains

$$\begin{aligned}\mathbf{b}_1&=\mathbf{S}_1\mathbf{a}_1\\ &=\mathbf{S}_1\mathbf{I}\mathbf{a}_1=\mathbf{S}_1\mathbf{A}\mathbf{a}_1+\mathbf{S}_1\mathbf{B}\mathbf{a}_1\\ &=\mathbf{S}_1\mathbf{A}\mathbf{S}_2\mathbf{a}_2+\mathbf{S}_1\mathbf{B}\mathbf{a}_1\\ &=\mathbf{S}_1\mathbf{A}\mathbf{S}_2\mathbf{A}\mathbf{a}_2+\mathbf{S}_1\mathbf{A}\mathbf{S}_2\mathbf{B}\mathbf{a}_2+\mathbf{S}_1\mathbf{B}\mathbf{a}_1.\end{aligned}\qquad(11)$$

Let

$$\mathbf{M}=\mathbf{S}_1\mathbf{A}\mathbf{S}_2\mathbf{A},\quad \mathbf{N}=\mathbf{S}_1\mathbf{A}\mathbf{S}_2.\qquad(12)$$

Then

$$\mathbf{b}_1=\mathbf{M}\mathbf{a}_2+\mathbf{N}(\mathbf{B}\mathbf{a}_2)+\mathbf{S}_1(\mathbf{B}\mathbf{a}_1).\qquad(13)$$

It should be noted that the last two terms on the

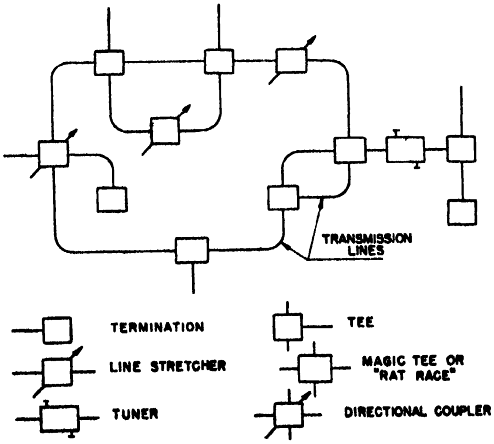


FIG. 1.

right side of (13) contain $\mathbf{Ba}_2$ and $\mathbf{Ba}_1$ as column vectors. From (7) and (8) it may be seen that $\mathbf{B}$ has all zeros in the central block. Hence if $\mathbf{B}$ operates on a column vector it replaces the $r$ rows whose indices run from $p-r+1$ through $p$ by zeros. Thus $\mathbf{B}$ has the effect of removing from $\mathbf{a}_1$ and $\mathbf{a}_2$ those elements which are unknown, namely, the wave amplitudes in the interconnecting transmission lines. The remaining elements of $\mathbf{a}_1$ and $\mathbf{a}_2$ are the incident waves in the transmission lines feeding the combined network and are assumed known.

The procedure which leads from (4) to (13) may be iterated to give the following sequence of equations:

$$\begin{aligned}
\mathbf{b}_1 &= \mathbf{Ma}_2+\mathbf{N}(\mathbf{Ba}_2)+\mathbf{S}_1(\mathbf{Ba}_1),\\
\mathbf{b}_1 &= \mathbf{M}^2\mathbf{a}_2+\mathbf{MN}(\mathbf{Ba}_2)+\mathbf{MS}_1(\mathbf{Ba}_1)\\
&\qquad +\mathbf{N}(\mathbf{Ba}_2)+\mathbf{S}_1(\mathbf{Ba}_1),\\
\mathbf{b}_1 &= \mathbf{M}^3\mathbf{a}_2+\mathbf{M}^2\mathbf{N}(\mathbf{Ba}_2)+\mathbf{M}^2\mathbf{S}_1(\mathbf{Ba}_1)\\
&\quad +\mathbf{MN}(\mathbf{Ba}_2)+\mathbf{MS}_1(\mathbf{Ba}_1)\\
&\qquad +\mathbf{N}(\mathbf{Ba}_2)+\mathbf{S}_1(\mathbf{Ba}_1),\\
&\ \ \vdots \qquad\qquad \vdots \qquad\qquad \vdots
\end{aligned}\tag{14}$$

Note that in this sequence of equations, with the exception of the 1st term, all terms on the right side have column vectors operated on by $\mathbf{B}$.

The Eq. (14) may be rearranged as follows:

$$\begin{aligned}
\mathbf{b}_1 &= \mathbf{Ma}_2+[\mathbf{N}(\mathbf{Ba}_2)+\mathbf{S}_1(\mathbf{Ba}_1)],\\
\mathbf{b}_1 &= \mathbf{M}^2\mathbf{a}_2+(\mathbf{M}+\mathbf{I})[\mathbf{N}(\mathbf{Ba}_2)+\mathbf{S}_1(\mathbf{Ba}_1)],\\
\mathbf{b}_1 &= \mathbf{M}^3\mathbf{a}_2+(\mathbf{M}^2+\mathbf{M}+\mathbf{I})[\mathbf{N}(\mathbf{Ba}_2)+\mathbf{S}_1(\mathbf{Ba}_1)].\\
&\ \ \vdots \qquad\qquad \vdots \qquad\qquad \vdots
\end{aligned}\tag{15}$$

Let the characteristic equation of the matrix $\mathbf{M}$ be

$$\sum_n C_n\lambda^n = 0. \tag{16}$$

It will be shown in the appendix that the term $C_0$ vanishes in (16) and furthermore that there can be no more than $r+1$ non-vanishing terms.

By Cayley-Hamilton's theorem the matrix $\mathbf{M}$ satisfies its characteristic equation and

$$\sum_n C_n\mathbf{M}^n = 0. \tag{17}$$

If possible, it is convenient to normalize the coefficients in (16) in such a way that

$$\sum C_n = 1. \tag{18}$$

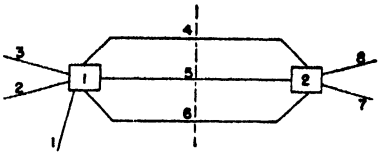


FIG. 2. $p=6$, $q=5$, and $r=3$.

If the sum of the coefficients of the characteristic equation vanishes, this normalization is impossible. This case will be discussed in the appendix.

Multiply Eqs. (15) by the coefficients $C_n$ and sum over the index $n$.

$$\sum_{n=1}^{r+1} C_n\mathbf{b}_1 = \left[\sum_{n=1}^{r+1} C_n\mathbf{M}^n\right]\mathbf{a}_2+\sum_{n=1}^{r+1} C_n\mathbf{S}_1(\mathbf{Ba}_1)$$
$$+\sum_{n=1}^{r+1} C_n\mathbf{N}(\mathbf{Ba}_2)+\sum_{n=2}^{r+1} C_n\mathbf{MN}(\mathbf{Ba}_2)+\cdots. \tag{19}$$

Substituting (17) and (18) in (19)

$$\begin{aligned}
\mathbf{b}_1 = [\mathbf{I}+(1-C_1)\mathbf{M}+(1-C_1-C_2)\mathbf{M}^2+\cdots]&\\
\times[\mathbf{N}(\mathbf{Ba}_2)+\mathbf{S}_1(\mathbf{Ba}_1)].&
\end{aligned}\tag{20}$$

It should be noted that $\mathbf{a}_1$ and $\mathbf{a}_2$ appear in (20) only in combination with $\mathbf{B}$. Equation (20) is part of the solution as it expresses the amplitudes of waves leaving the junction number (1) as a linear combination of the amplitudes of waves entering the combined network.

Since

$$\mathbf{S}_1\mathbf{Ba}_2 = 0 \tag{21}$$

and

$$\mathbf{NBa}_1 = 0,$$

Eq. (20) can be written as

$$\begin{aligned}
\mathbf{b}_1 = [\mathbf{I}+(1-C_1)\mathbf{M}+(1-C_1-C_2)\mathbf{M}^2+\cdots]&\\
\times(\mathbf{N}+\mathbf{S}_1)\mathbf{B}(\mathbf{a}_1+\mathbf{a}_2).&
\end{aligned}\tag{22}$$

The amplitudes of waves leaving the junction number (2) remain to be calculated. This is easily done making use of (22). From (5) and (10)

$$\mathbf{b}_2 = \mathbf{S}_2\mathbf{a}_2 = \mathbf{S}_2(\mathbf{A}+\mathbf{B})\mathbf{a}_2 = \mathbf{S}_2\mathbf{Ab}_1+\mathbf{S}_2\mathbf{Ba}_2. \tag{23}$$

Substitute (22) in (23)

$$\begin{aligned}
\mathbf{b}_2 = \mathbf{S}_2\mathbf{A}[\mathbf{I}+(1-C_1)\mathbf{M}+(1-C_1-C_2)\mathbf{M}^2+\cdots]&\\
\times(\mathbf{N}+\mathbf{S}_1)\mathbf{B}(\mathbf{a}_1+\mathbf{a}_2)+\mathbf{S}_2\mathbf{Ba}_2,&
\end{aligned}\tag{24}$$

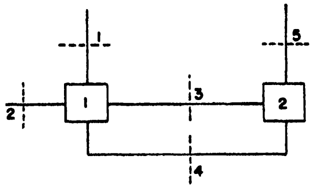


FIG. 3.

which may be written as

$$\mathbf{b}_2=\{\mathbf{S}_2+\mathbf{S}_2\mathbf{A}[\mathbf{I}+(1-C_1)\mathbf{M}+\cdots](\mathbf{N}+\mathbf{S}_1)\}\times\mathbf{B}(\mathbf{a}_1+a_2).\quad(25)$$

Equations (22) and (25) may be combined to form the scattering equation for the combined network.

$$\mathbf{B}(\mathbf{b}_1+\mathbf{b}_2)=\mathbf{S}\mathbf{B}(\mathbf{a}_1+\mathbf{a}_2)\quad(26)$$

where

$$\mathbf{S}=\mathbf{B}\{\mathbf{S}_2+(\mathbf{I}+\mathbf{S}_2\mathbf{A})[\mathbf{I}+(1-C_1)\mathbf{M}+(1-C_1-C_2)\mathbf{M}^2+\cdots](\mathbf{N}+\mathbf{S}_1)\}\mathbf{B}.\quad(27)$$

**S** as given by (27) is the scattering matrix of the combined network.

The network scattering matrix **S** has the following form:

$$\mathbf{S}=\left[\begin{array}{c|c|c} & 0 & \\ \hline 0 & 0 & 0 \\ \hline & 0 & \\ \underbrace{\hphantom{000}}_{p-r} & \underbrace{\hphantom{0}}_{r} & \underbrace{\hphantom{000}}_{q-r} \end{array}\right]\begin{matrix}\}p-r\\ \}r\\ \}q-r\\ \phantom{0}\end{matrix}.$$

The $r$ rows and columns which are zero may be removed to form a matrix of order $p+q-2r$. In a similar way the column vectors $\mathbf{B}(\mathbf{b}_1+\mathbf{b}_2)$ and $\mathbf{B}(\mathbf{a}_1+\mathbf{a}_2)$ have zeros in $r$ rows beginning at the $p-r+1$ row. These zeros may be removed to form $p+q-2r$ rowed vectors. This process of removing zeros reduces the matrices to a dimension equal to the number of transmission lines feeding the combined network. The rows and columns removed are the ones associated with the interconnecting transmission lines.

For some purposes it is desirable to know the amplitudes of waves in the interconnecting transmission lines in terms of amplitudes of waves incident upon the network. These amplitudes are contained in $\mathbf{b}_1$ and $\mathbf{b}_2$ and may be expressed as

$$\mathbf{A}\mathbf{b}_1=\mathbf{A}[\mathbf{I}+(1-C_1)\mathbf{M}+(1-C_1-C_2)\mathbf{M}^2+\cdots]\times(\mathbf{N}+\mathbf{S}_1)\mathbf{B}(\mathbf{a}_1+\mathbf{a}_2)\quad(28)$$

and

$$\mathbf{A}\mathbf{b}_2=\mathbf{A}\{\mathbf{S}_2+\mathbf{S}_2\mathbf{A}[\mathbf{I}+(1-C_1)\mathbf{M}+\cdots]\times(\mathbf{N}+\mathbf{S}_1)\}\mathbf{B}(\mathbf{a}_1+\mathbf{a}_2).\quad(29)$$

The left-hand side of Eqs. (28) and (29) consists only of the amplitudes of waves in interconnecting transmission lines of the network.

To recapitulate, Eqs. (26)–(29) express the unknown complex wave amplitudes of waves in the transmission lines of a microwave network of the type illustrated in Fig. 2 as functions of the known amplitudes. In particular the left side of (26) consists of the amplitudes of waves leaving the network (along the lines 1, 2, 3, 7, and 8 of Fig. 2) expressed as a linear combination of the complex amplitudes, assumed known, of waves entering the network. In a similar way Eqs. (28) and (29) express the complex amplitudes of waves in the interconnecting transmission lines (4, 5, and 6, Fig. 2) as linear combinations of the known complex amplitudes of waves entering the network (along lines 1, 2, 3, 7, and 8 of Fig. 2).

## EXAMPLE

As an example of the application of the formalism, consider the microwave network illustrated schematically in Fig. 3.

The "magic tee" has the property of dividing the power incident upon it in any one of its four feeding transmission lines equally between the two adjacent lines. For example, if a wave is sent in on line 2, it leaves the junction No. 1 *via* lines 1 and 4. Of course, the wave which leaves junction No. 1 *via* line 4 returns again after interacting with junction No. 2. This leads, in turn, to a new set of waves scattered by junction No. 1 and so on. By locating the terminals in (1), (2), (3), or (4) properly, the scattering matrix of the junction No. 1 may be written as

$$\mathbf{S}_1=\frac{1}{\sqrt{2}}\left[\begin{array}{cc|cc|c} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right].\quad(30)$$

As an example of the meaning of $\mathbf{S}_1$ note that the third column of $\mathbf{S}_1$ indicates that a wave of amplitude 1 incident on the junction in transmission line 3 leaves the junction along lines 1 and 4 with amplitudes $1/\sqrt{2}$ and $-1/\sqrt{2}$, respectively. The phases of these waves are to be taken at their respective reference planes.

In a similar way the junction No. 2 is a matched symmetrical tee. Its scattering matrix is assumed to be

$$\mathbf{S}_2=\left[\begin{array}{c|cc|c} 0 & 0 & & 0 \\ \hline 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ & 1/\sqrt{2} & \frac{1}{2} & -\frac{1}{2} \\ \hline 0 & 1/\sqrt{2} & -\frac{1}{2} & \frac{1}{2} \end{array}\right]. \quad (31)$$

It is possible to write the scattering matrix of any matched symmetrical tee in this way merely by locating the terminals correctly. However, the terminals 3 (or 4) have already been fixed in position. Therefore, it is necessary to assume that either the line 3 or the line 4 has been adjusted to have the correct length.

It may be noted that both $\mathbf{S}_1$ and $\mathbf{S}_2$ are symmetric and unitary. These are general properties and are characteristic of all scattering matrices.

Turning now to the two auxiliary matrices **A** and **B**, these have the following forms:

$$\mathbf{A}=\left[\begin{array}{c|cc|c} 0 & 0 & & 0 \\ \hline 0 & 1 & 0 & 0 \\ & 0 & 1 & \\ \hline 0 & 0 & & 0 \end{array}\right], \quad (32)$$

$$\mathbf{B}=\left[\begin{array}{cc|c|c} 1 & 0 & 0 & 0 \\ 0 & 1 & & \\ \hline 0 & & 0 & 0 \\ \hline 0 & & 0 & 1 \end{array}\right]. \quad (33)$$

The two matrices **M** and **N** may now be computed (see Eq. (12)).

$\mathbf{N}=\mathbf{S}_1\mathbf{A}\mathbf{S}_2$

$$=\frac{1}{2\sqrt{2}}\left[\begin{array}{c|cc|c} 0 & 0 & \sqrt{2} & \sqrt{2} \\ & \sqrt{2} & 1 & -1 \\ \hline 0 & -\sqrt{2} & -1 & +1 \\ & 0 & -\sqrt{2} & -\sqrt{2} \\ \hline 0 & 0 & & 0 \end{array}\right], \quad (34)$$

$\mathbf{M}=\mathbf{S}_1\mathbf{A}\mathbf{S}_2\mathbf{A}$

$$=\frac{1}{2\sqrt{2}}\left[\begin{array}{c|cc|c} 0 & 0 & \sqrt{2} & 0 \\ & \sqrt{2} & 1 & \\ \hline 0 & -\sqrt{2} & -1 & 0 \\ & 0 & -\sqrt{2} & \\ \hline 0 & 0 & & 0 \end{array}\right]. \quad (35)$$

Equation (16) may be obtained by using the method introduced in the appendix [see Eq. (45)]

$$0=\sum_n C_n\lambda^n=\lambda\det(\mathbf{R}-\mathbf{I}\lambda)=\lambda\begin{vmatrix} -(\lambda+\frac{1}{2}) & \dfrac{-1}{2\sqrt{2}} \\ 0 & -(\lambda+\frac{1}{2}) \end{vmatrix}$$

$$=\lambda(\lambda+\tfrac{1}{2})^2=\lambda^3+\lambda^2+\tfrac{1}{4}\lambda. \quad (36)$$

The coefficients $C_n$ must be normalized to satisfy Eq. (18). They then have the values

$$C_1=\frac{1}{9},\quad C_2=\frac{4}{9},\quad C_3=\frac{4}{9}. \quad (37)$$

The scattering matrix of the combined network may be computed by substituting in Eq. (27)

$$\mathbf{S}=\mathbf{B}\left\{\mathbf{S}_2+(\mathbf{I}+\mathbf{S}_2\mathbf{A})\left[\mathbf{I}+\frac{8}{9}\mathbf{M}+\frac{5}{9}\mathbf{M}^2\right](\mathbf{N}+\mathbf{S}_1)\right\}\mathbf{B}$$

$$=\frac{1}{3}\left[\begin{array}{cc|c|c} 0 & 2\sqrt{2} & 0 & 1 \\ 2\sqrt{2} & \frac{1}{3} & & -\dfrac{2\sqrt{2}}{3} \\ \hline 0 & & 0 & 0 \\ \hline 1 & -\dfrac{2\sqrt{2}}{3} & 0 & \dfrac{8}{3} \end{array}\right]. \quad (38)$$

If the two rows and columns which vanish are removed, the matrix becomes

$$\mathbf{S}=\frac{1}{3}\begin{bmatrix} 0 & 2\sqrt{2} & 1 \\ 2\sqrt{2} & \frac{1}{3} & -\dfrac{2\sqrt{2}}{3} \\ 1 & -\dfrac{2\sqrt{2}}{3} & \dfrac{8}{3} \end{bmatrix}. \quad (39)$$

It should be noted that this matrix is symmetric and unitary as it should be.

The matrices which can be used to calculate the amplitudes of waves in the connecting lines are obtained from Eqs. (28) and (29).

$$\mathbf{S}_c=\mathbf{A}\left[\mathbf{I}+\frac{8}{9}\mathbf{M}+\frac{4}{9}\mathbf{M}^2\right](\mathbf{N}+\mathbf{S}_1)\mathbf{B}$$

$$=\frac{1}{3}\left[\begin{array}{cc|c|c} \multicolumn{2}{c|}{0} & 0 & 0 \\ \hline \sqrt{2} & -\frac{1}{3} & & \frac{2\sqrt{2}}{3} \\ & & 0 & \\ 0 & \sqrt{2} & & -1 \\ \hline \multicolumn{2}{c|}{0} & 0 & 0 \end{array}\right]$$

$$\mathbf{S}_c'=\mathbf{A}\left\{\mathbf{S}_2+\mathbf{S}_2\mathbf{A}\left[\mathbf{I}+\frac{8}{9}\mathbf{M}+\frac{4}{9}\mathbf{M}^2\right](\mathbf{N}+\mathbf{S}_1)\right\}\mathbf{B}$$

$$=\frac{1}{9}\left[\begin{array}{cc|c|c} \multicolumn{2}{c|}{0} & 0 & 0 \\ \hline 0 & 3 & & 3\sqrt{2} \\ & & 0 & \\ 3 & \sqrt{2} & & -4 \\ \hline \multicolumn{2}{c|}{0} & 0 & 0 \end{array}\right]. \quad (41)$$

To illustrate the meaning of the matrices (38), (40), and (41), assume that a wave of unit amplitude is incident upon the combined network at terminals No. 1. The scattering matrix **S**, Eq. (39), states that the amplitudes of the waves leaving the network *via* lines (2) and (5) are $\frac{2}{3}\sqrt{2}$ and $\frac{1}{3}$, respectively. The matrix (40) states that there is a wave of amplitude $\sqrt{2}/3$ running in line (3) from junction No. 1 to No. 2. There is no wave moving in this direction along line (4). The matrix (41) states that there is no wave moving from junction No. 2 to No. 1 along (3), but that there is a wave of amplitude $\frac{1}{3}$ moving from right to left in line (4).

## APPENDIX

From Eqs. (3), (6), and (12) it is apparent that **M** has the form

$$\mathbf{M}=\left[\begin{array}{c|c|c} 0 & - & 0 \\ 0 & \mathbf{R} & 0 \\ 0 & 0 & 0 \\ p-r & r & q-r \end{array}\right]\begin{array}{l} \}p-r \\ \}r \\ \}q-r \end{array}. \quad (42)$$

The characteristic equation of **M** is defined as

$$\det(\mathbf{M}-\lambda\mathbf{I})=0. \quad (43)$$

Denote the middle square of **M** as a sub-matrix **R**. It is evident that

$$\det(\mathbf{M}-\lambda\mathbf{I})=\lambda^{q+p-2r}\det(\mathbf{R}-\mathbf{I}\lambda), \quad (44)$$

where the determinant on the right side of (32) is of order $r$. In order to reduce the characteristic polynomial to its lowest degree, Eq. (16) is defined by the relation

$$0=\sum_n C_n\lambda^n=\lambda\det(\mathbf{R}-\mathbf{I}\lambda). \quad (45)$$

Thus the characteristic polynomial is of degree $r+1$ and the index $n$ runs from 1 to $r+1$.

It will be remembered that the above described computational method depended for its validity on the non-vanishing of the sum of the coefficients of (33). A necessary and sufficient condition for the vanishing of this sum is the occurrence of

$$\lambda=1$$

as an eigenvalue of the matrix **M**. The values of the eigenvalues of **M** depend upon the scattering matrices $\mathbf{S}_1$ and $\mathbf{S}_2$. However, the locations of the reference planes or terminals in the interconnecting transmission lines are arbitrary. By changing their locations an accidental occurence of the root $\lambda=1$ can be eliminated.

# The Optics of Three-Electrode Electron Guns*

S. G. Ellis
*The McLennan Laboratory, Department of Physics, University of Toronto, Toronto, Canada*

(Received April 17, 1947)

The paper is concerned primarily with the electron guns employed in electron microscopy. The position of the image of the cathode, the magnification, and the divergence of the beam leaving the anode are calculated for a three-electrode electron gun with plane electrodes and circular apertures both for zero and battery bias. The simplifying assumption is made that the electrostatic lenses are thin. The results are compared with the experiments of Johannson. The extension of the results to the gun used in the transmission electron microscope is considered with the aid of electrolytic trough measurements. The case of a gun biased by a cathode resistor is also discussed.

## INTRODUCTION

THE illuminating system of the transmission electron microscope was first treated analytically by von Borries and Ruska.[1] They considered the electron gun as a source of electrons of fixed cross section and position and considered the dependence of the semi-angle of the illuminating beam at the specimen on the strength of the condenser lens. Marton and Hutter[2] extended this treatment to include the converging effect on the illuminating beam of that part of the objective field which lies between the specimen and the condenser lens. Finally, Hillier and Baker[3] have considered, semi-empirically, the influence of the filament height on the angular aperture of the illumination by treating the electron gun as a converging lens containing a physical aperture which may itself limit the aperture of the illumination.

Hitherto the electron gun has received either very general physical treatment,[4] or rather special semi-empirical treatment[5] in the literature. Since these treatments do not greatly clarify the behavior of the particular types of electron gun employed in electron microscopy, an attempt has been made in this paper to develop an analytic theory of a very simplified electron gun. The main simplifications are the treatment of the electrodes as plane, and the variations of axial potential as linear. The latter simplification is equivalent to treating the lenses as thin lenses. Later, this theory is used, in conjunction with electrolytic trough measurements, to describe the behavior of the guns employed in practice.

This analytic treatment is aided by a theorem attributed to Schlesinger[6] which may be called the "Transit Time Theorem." Consider an electron leaving an object (cathode or image) at $A$, Fig. 1, passing through a lens at $B$ and being focused to $C$. An axis $Oz$ is taken, coaxial with the lens, positive in the direction of electron motion and with origin at the lens. Take the zero of time as the epoch at which the electron passes through the lens $OB$. $T_0$ is the epoch at which the electron leaves $A$ and $T_I$ the epoch at which it arrives at $C$. It is supposed that the lens is thin so that radial forces act on the electron only while it is passing the $OB$ plane. If then a radial impulse $I$, positive if centrifugal, acts on the electron when passing through the lens at a distance $R$ from the center, then Schlesinger shows that the condition for image formation is:

$$I \propto R \tag{1}$$

and that under these conditions

$$(1/T_I)-(1/T_0)=-(I/mR) \tag{2}$$

$$=1/T_F \text{ say.} \tag{3}$$

In Eq. (2) $m$ is the mass of the electron. Equation (2) immediately leads to Gaussian dioptrics in the magnetic lens and the bipotential lens,

---

* Based on Section A of a thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy in the University of Toronto. Part of this work was supported by the grant of a Fellowship by the National Research Council, Ottawa, Canada.

[1] B. von Borries and E. Ruska, Zeits. f. tech. Physik. **20**, 225 (1939).

[2] L. Marton and R. C. E. Hutter, Phys. Rev. **65**, 161 (1944).

[3] J. Hillier and R. F. Baker, J. App. Phys. **16**, 469 (1945).

[4] L. M. Field, Rev. Mod. Phys. **18**, 353 (1946).

[5] G. A. Morton, Rev. Mod. Phys. **18**, 362 (1946).

[6] K. Schlesinger, Proc. I.R.E. **32**, 483 (1944).

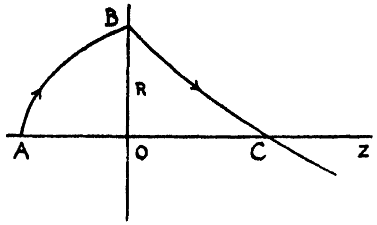


FIG. 1. General electron trajectory through a thin lens.

where the axial velocity of the electron is constant outside the lens and the trajectories are linear there. In the immersion objective the dioptrics will not be Gaussian, both because the axial velocity is continually changing and also because the strength of the lens (proportional to $I/R$) will depend on the object distance $OA$. The magnification, $M$, for the thin lens is, quite generally, given by

$$M = T_I/T_0. \tag{4}$$

## SIMPLIFYING ASSUMPTIONS AND DEFINITIONS

The filament or cathode is considered as a plane circular emitting area of radius $r_e$ (see Fig. 2). The shield, or grid, is considered as a thin sheet with a central aperture of radius $r_k''$, distant $d_2$ from a plane anode. An axis $O_s z$ with origin $O_s$ at the center of the shield aperture and positive in the direction of electron motion is chosen to coincide with the axis of symmetry of the electron gun. The assumed variation of axial potential is shown in Fig. 3.

The following factors have now to be taken into consideration. The electrons leave the filament with a velocity distribution both in magnitude and direction governed by the temperature of the filament. The electrostatic field in the vicinity of the shield exerts a convergent effect upon the electrons. The electron speeds are nowhere constant within the electron gun, and it is therefore convenient to use the transit time method of Schlesinger taking zero time as the instant at which the electron passes through the shield.

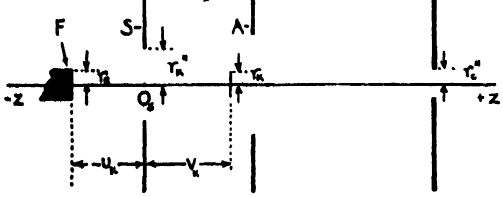


FIG. 2. Symbols used in text. $F$ represents the filament, $S$ the shield, and $A$ the anode.

To commence, consider an electron which leaves the filament at the axis with zero axial velocity and with transverse thermal velocity $v_T$ cm per sec. (or $V_T$ in electron volts). It describes a parabolic path (Fig. 4) until it reaches the shield aperture at a distance $R$ from the axis. At this point it receives a centripetal impulse $I_k$ and returns to the axis at a distance $v_k$ from the shield, following another parabola until it reaches the anode aperture where it receives a centrifugal impulse $I_A$ traveling thereafter in a straight line towards the condenser lens of the microscope.

## NOTATION

Where possible the notation employed by Hillier and Baker[3] has been used in this section.

$r_e$ = radius of emitting area, assumed plane
$r_k''$ = radius of hole in shield
$r_c''$ = radius of condenser aperture
$r_0'$ = radius of area of specimen plane illuminated by electrons
$V_k$ = potential at origin $O_s$
$V_k'$ = potential of shield
$V$ = potential of anode
$V_T$ = transverse (thermal) velocity of electron leaving filament in electron volts
$E_1$ = electrostatic field strength between filament and shield = $V_k/u_k$
$E_2$ = electrostatic field strength between shield and anode = $-V/d_2$ (if $V_k \ll V$)
$v_T$ = transverse (thermal) velocity of electron leaving filament in cm per sec.
$T_0$ = epoch at which electron leaves filament (negative with sign convention used here)
$T_I$ = epoch at which electron arrives at real image formed by cathode lens
$T_{SA}$ = transit time of electron from shield to anode
$T_0'$ = epoch at which electron leaves real image formed by cathode lens
$T_I'$ = epoch at which electron 'passes' virtual image formed by anode aperture
$u_k$ = distance from shield to filament (negative on above sign convention) commonly called the filament height
$v_k$ = distance from shield to image formed by cathode lens
$d_1$ = distance from shield to condenser aperture
$d_2$ = distance from shield to anode
$R$ = radial displacement of electron at shield
$\dot{R}$ = radial velocity of electron between shield and anode

$R_A$ = radial displacement of electron at anode
$I_k = I_k(R)$ = centripetal impulse acting on electron on passing shield
$I_A = I_A(R_A)$ = centrifugal impulse acting on electron on passing anode
$\alpha_c$ max. = maximum angular aperture of illumination at specimen
$\beta$ = angular aperture of beam at cathode side of anode
$\beta'$ = angular aperture of beam leaving the electron gun
$\gamma$ = angle between $O_sz$ and center of beam from off-axial point of cathode, see Fig. 8
$M_k$ = magnification produced by cathode lens
$M_A$ = magnification produced by anode lens
$M_G$ = over-all magnification produced by electron gun
$M_c$ = magnification produced by condenser lens
$e$ = charge of electron
$m$ = mass of electron.

## THEORY

From the transit time theory of Schlesinger

$$(1/T_I)-(1/T_0)=-(I_k/mR)=1/T_F. \qquad (5)$$

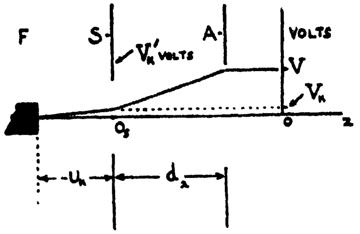


FIG. 3. Potential distribution assumed.

Further

$$-u_k=\frac{1}{2}\left(\frac{eV_k}{mu_k}\right)T_0^2. \qquad (6)$$

At the plane of the shield the electron will have gained an axial component of velocity $(-2eV_k/m)^{1/2}$. Hence

$$v_k=\left(-\frac{2eV_k}{m}\right)^{1/2}T_I-\frac{1}{2}\frac{eV}{md_2}\cdot T_I^2 \qquad (7)$$

and

$$T_I=\frac{2(-Ve/2md_2)^{1/2}\cdot[\lambda(1-\epsilon)]^{1/2}\pm\{4(-Ve/2md_2)\cdot[\lambda(1-\epsilon)]+4(-Ve/2md_2)v_k\}^{1/2}}{-2(-Ve/2md_2)}, \qquad (8)$$

where

$$\lambda=u_kr_k''/(\pi u_k-r_k'') \qquad (9)$$

and

$$\epsilon=-\pi(d_2/r_k'')\cdot(V_k'/V). \qquad (10)$$

Then with an obvious choice of sign

$$T_I=\{[\lambda(1-\epsilon)+v_k]^{1/2}-[\lambda(1-\epsilon)]^{1/2}\}/(-Ve/2md_2)^{1/2}. \qquad (11)$$

To obtain the value of $I_k/mR$ notice that if the lens may be treated as a thin lens[7]

$$\frac{I_k}{mR}=-\left(-\frac{2V_ke}{m}\right)^{1/2}\cdot\left[\frac{E_2-E_1}{4V_k}\right], \qquad (12)$$

but[8]

$$V_k=V\lambda(1-\epsilon)/d_2 \qquad (13)$$

and, therefore,

$$\frac{I_k}{mR}=\left[-\frac{Ve}{2md_2}\right]^{1/2}\cdot\left[\frac{1}{2[\lambda(1-\epsilon)]^{1/2}}+\frac{[\lambda(1-\epsilon)]^{1/2}}{2u_k}\right]. \qquad (14)$$

From Eqs. (6), (11), and (14)

$$\frac{1}{T_0}=\left(-\frac{Ve}{2md_2}\right)^{1/2}\cdot\frac{1}{u_k}\cdot[\lambda(1-\epsilon)]^{1/2} \qquad (15)$$

$$\frac{1}{T_I}=\left(-\frac{Ve}{2md_2}\right)^{1/2}\cdot\frac{1}{[\lambda(1-\epsilon)+v_k]^{1/2}-[\lambda(1-\epsilon)]^{1/2}} \qquad (16)$$

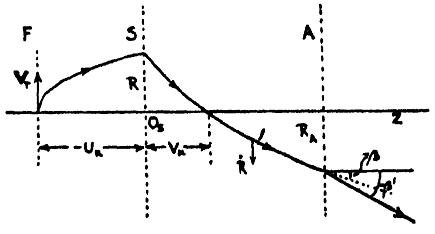


FIG. 4. An electron trajectory for the three-electrode electron gun.

---

[7] Zworykin, Morton, Ramberg, Hillier, and Vance, *Electron Optics and the Electron Microscope* (John Wiley and Sons, Inc., New York, 1945), p. 444.

[8] Zworykin, Morton, Ramberg, Hillier, and Vance, *Electron Optics on the Electron Microscope* (John Wiley and Sons, Inc., New York, 1945), pp. 383–385, particularly equation 11.95.

$$\frac{1}{T_F}=\left(-\frac{Ve}{2md_2}\right)^{\frac{1}{2}}$$

$$\cdot\left\{[\lambda(1-\epsilon)]^{\frac{1}{2}}\left[\frac{1}{2\lambda(1-\epsilon)}+\frac{1}{2u_k}\right]\right\}. \quad (17)$$

Substituting from Eqs. (15) to (17) into Eq. (5) and solving for $v_k$

$$v_k=4[\lambda(1-\epsilon)]\left\{\frac{1+(1+3[\lambda(1-\epsilon)]/u_k)}{(1+3[\lambda(1-\epsilon)]/u_k)^2}\right\}. \quad (18)$$

Values of $v_k$ are plotted in Fig. 5 for typical values of the other variables. The non-Gaussian dioptrics of the system is apparent from Eq. (18).

### DIVERGENCE OF THE BEAM

One of the factors controlling the intensity of illumination in the electron microscope will be the divergence of the electron beam leaving the electron gun, for this will determine the percentage of electrons leaving the electron gun which pass through the condenser diaphragm.

If $\beta$ is the angle to the axis $O_sz$ at which the peripheral beam approaches the anode then, from Fig. 4

$$\tan\beta=\frac{\dot{R}}{(-2eV/m)^{\frac{1}{2}}}=\frac{I_k/m+\mathbf{v}_T}{(-2eV/m)^{\frac{1}{2}}}, \quad (19)$$

but

$$I_k/m=-R/T_F \quad (20)$$

and

$$R=-\mathbf{v}_T T_0; \quad (21)$$

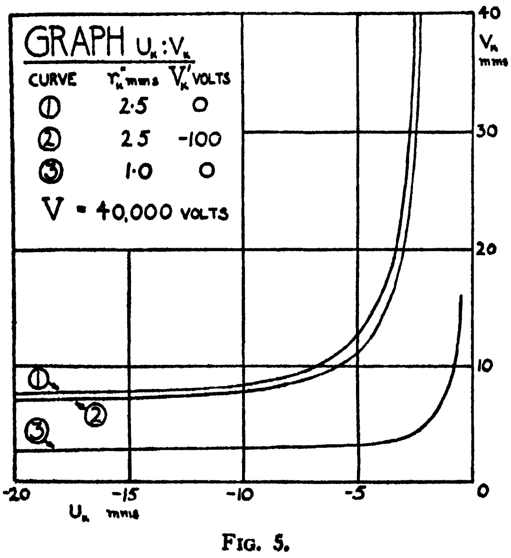


FIG. 5.

therefore

$$\tan\beta=\left(\frac{V_T}{V}\right)^{\frac{1}{2}}\left[1+\frac{T_0}{T_F}\right]. \quad (22)$$

Substituting in Eq. (22) from Eqs. (15) and (17)

$$\tan\beta=\frac{3}{2}\left(\frac{V_T}{V}\right)^{\frac{1}{2}}\left[1-\frac{1}{3(1-\epsilon)}+\frac{\pi u_k}{3r_k''(1-\epsilon)}\right]. \quad (23)$$

The further divergence produced in the electron beam by the change in field strength at the anode aperture may now be determined. Then

$$\tan\beta'=\frac{\dot{R}+I_A/m}{(-2eV/m)^{\frac{1}{2}}} \quad (24)$$

$$=\tan\beta+I_A/m(-2eV/m)^{\frac{1}{2}}. \quad (25)$$

But,[7]

$$\frac{I_A}{mR_A}=\left(-\frac{2eV}{m}\right)^{\frac{1}{2}}\cdot\frac{1}{4d_2} \quad (26)$$

$$\frac{I_A}{m(-2eV/m)^{\frac{1}{2}}}=\frac{R_A}{4d_2} \quad (27)$$

and

$$\tan\beta'=\tan\beta+R_A/4d_2. \quad (28)$$

Further

$$R+R_A=\dot{R}T_{SA} \quad (29)$$

and

$$T_{SA}=T_I \text{ (for } v_k=d_2), \quad (30)$$

that is

$$T_{SA}=(-2d_2m/Ve)^{\frac{1}{2}}\{[\lambda(1-\epsilon)+d_2]^{\frac{1}{2}}-[\lambda(1-\epsilon)]^{\frac{1}{2}}\} \quad (31)$$

and approximating

$$T_{SA}=(-2d_2m/Ve)^{\frac{1}{2}}\cdot d_2^{\frac{1}{2}} \quad (32)$$

giving

$$R_A=2d_2\tan\beta+2\left(\frac{V_T}{V}\right)^{\frac{1}{2}}\cdot\frac{d_2^{\frac{1}{2}}u_k}{[\lambda(1-\epsilon)]^{\frac{1}{2}}} \quad (33)$$

and

$$\tan\beta'=\frac{3}{2}\tan\beta+\frac{1}{2}\left(\frac{V_T}{V}\right)^{\frac{1}{2}}\frac{u_k}{[\lambda d_2(1-\epsilon)]^{\frac{1}{2}}}. \quad (34)$$

The negative values of $\beta'$ obtained from Eq. (34) indicate that the electrons from the filament cross the axis of the gun. Values of $\tan\beta'$ are plotted in Fig. 6. It has been assumed here that $R$ is less than the radius of the $-V_T$ equipotential surfaces at the cathode shield, which is true for the range of $u_k$ plotted in the graphs.

## MAGNIFICATION

It would appear from Fig. 5 that for the range of filament heights commonly employed in the electron microscope, the cathode lens forms a real image of the filament between the shield and the anode (i.e., $v_k < d_2$). First, therefore, the magnification $M_k$ produced by the cathode lens may be calculated.

By the transit time method of Schlesinger

$$M_k = T_I / T_0. \tag{35}$$

Negative values obtained for $M_k$ indicate inversion of the image with respect to the object. Using Eqs. (15) and (16)

$$M_k = [\lambda(1-\epsilon)]^{\frac{1}{2}}\{[\lambda(1-\epsilon)+v_k]^{\frac{1}{2}} - [\lambda(1-\epsilon)]^{\frac{1}{2}}\}/u_k. \tag{36}$$

The anode lens forms a virtual image of the real image mentioned above. For $v_k < d_2$ this virtual image lies between $v_k$ and $d_2$ and is demagnified. If $M_A$ is the magnification produced by the anode lens alone then

$$M_A = T_I' / T_0' \tag{37}$$

but

$$(1/T_I') - (1/T_0') = -(I_A/mR_A) \tag{38}$$

$$= 1/T_F' \text{ say} \tag{39}$$

so that

$$M_A = T_F'/(T_0' + T_F') \tag{40}$$

writing

$$T_0' = -(T_{SA} - T_I) \tag{41}$$

and using Eqs. (31) and (16)

$$M_A = \frac{2d_2^{\frac{1}{2}}\{[\lambda(1-\epsilon)+v_k]^{\frac{1}{2}} + [\lambda(1-\epsilon)+d_2]^{\frac{1}{2}}\}}{(d_2 - v_k) + 2d_2^{\frac{1}{2}}\{[\lambda(1-\epsilon)+v_k]^{\frac{1}{2}} + [\lambda(1-\epsilon)+d_2]^{\frac{1}{2}}\}}. \tag{42}$$

Values of $M_k$ and $M_A$ are plotted against $u_k$ in Fig. 7.

## AREA OF SPECIMEN ILLUMINATED IN THE ELECTRON MICROSCOPE

On the basis of the foregoing theory one might expect the radius of the illuminated part of the specimen in an electron microscope $r_0'$ to be given by

$$r_0' = r_e M_k M_A M_c, \tag{43}$$

where $M_c$ is the magnification produced by the condenser lens. Writing

$$M_k M_A = M_G \tag{44}$$

the magnification produced by the electron gun, and taking $M_c = 1$ then

$$r_0' = r_e M_G. \tag{45}$$

Equation (45) applies only to the case in which the virtual image of the cathode due to the anode aperture is focused on to the specimen by the condenser lens.

For values of $u_k$ that make $\tan\beta'$ small, another factor may have a decisive effect on the value of $r_0'$. The condenser aperture may act as a field limiting aperture for the electron gun even though the image produced by the gun is not in the plane of the condenser aperture.

It will be seen from Fig. 8 that electrons leaving a point in the cathode distant $r_e'$ from the axis leave the electron gun in a cone of semi-angle $\beta'$, the axis of the cone making an angle $\gamma$ with the axis of the system. If the radius of the condenser lens diaphragm is $r_c''$, and it is situated a distance $d_1$ from $O_s$, the condition that electrons from $P$ shall pass the condenser lens is then, approximately

$$(\gamma - \beta')(d_1 - v_k) \leqslant r_c'' - |M_G| r_e' \tag{46}$$

with

$$r_e' \leqslant r_c. \tag{47}$$

It is desirable, therefore, to investigate the dependence of $\gamma$ upon $r_e'$. Considering the ray $Pcc'$ and employing the methods of the preceding sections, it can be shown that

$$\gamma = \frac{r_e'}{4}\left\{\left[\frac{\lambda(1-\epsilon)}{d_2}\right]^{\frac{1}{2}}\left[\frac{1}{\lambda(1-\epsilon)} + \frac{1}{u_k}\right] + \frac{1}{d_2}\right\}. \tag{48}$$

Neglecting the term $|M_G| r_e'$ in Eq. (46), and taking

$$r_c''/(d_1 - v_k) = \alpha_c \max, \tag{49}$$

we can calculate the radius $r_e''$ of the area of the cathode from which electrons can pass the condenser aperture. Equation (45) presumes that the electron gun and the specimen are equidistant from the condenser lens. Then from Eqs. (46),

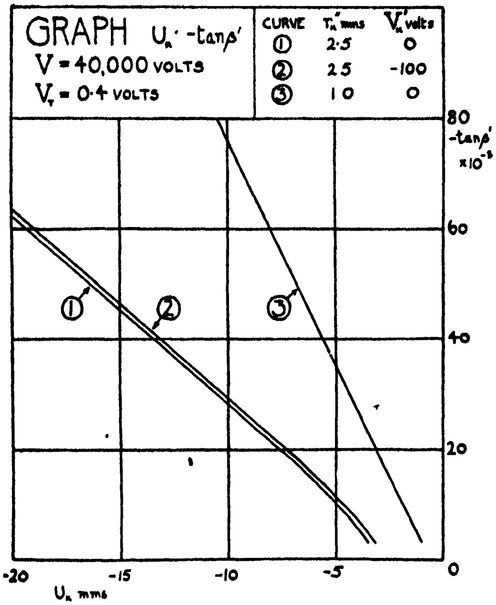


FIG. 6.

(48), and (49)

$$r_e'' = 4(\alpha_c \max + |\beta'|) \Big/ \left\{ \left[\frac{\lambda(1-\epsilon)}{d_2}\right]^{\frac{1}{2}} \times \left[\frac{1}{\lambda(1-\epsilon)} + \frac{1}{u_k}\right] + \frac{1}{d_2} \right\} \quad (50)$$

$$r_0' = r_e'' M_G. \quad (51)$$

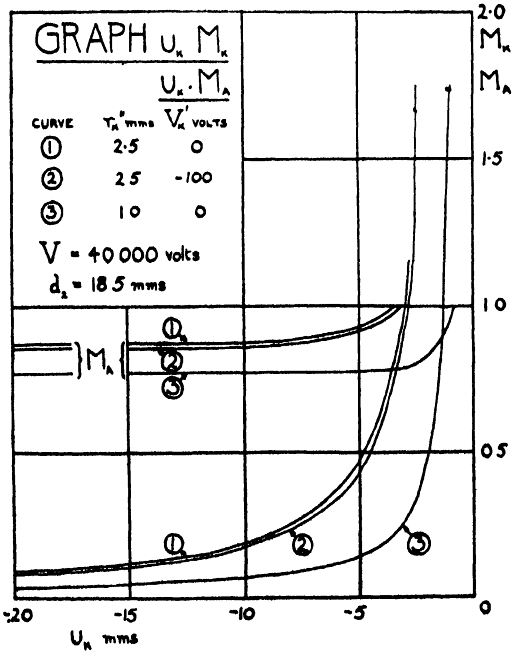


FIG. 7.

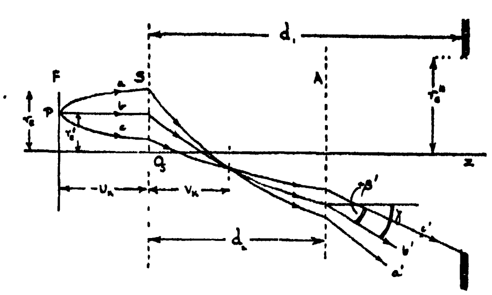


FIG. 8. Trajectories in beam from off-axial point in the filament.

In practice $r_0'$ will be given by the smaller of the values in Eqs. (45) or (51). In Fig. 9 $r_0'$ is plotted against $u_k$ the full line showing the variation from Eq. (45), and the broken line the variation from Eq. (51) for a sufficiently large cathode that relation 47 is maintained at all $u_k$. (The corresponding curves for conditions 1 and 2 lie off the graph with the scale chosen).

It has sometimes been stated that the smallness of the effective source provided by the electron gun is due to the presence of a cross-over of smaller radius than the image of the cathode. From Fig. 10 it can be seen that such a cross-over exists only if $\beta' < \gamma$. This is approximately the condition for Eq. (51) to limit the area of illumination of the specimen.

On the basis of the foregoing calculation one might conclude that the source in an electron microscope operated with a zero-biased, or battery-biased gun and with the filament at a normal height is a reduced image of the filament and not a cross-over. This will be discussed later.

One other point should be mentioned. Only the

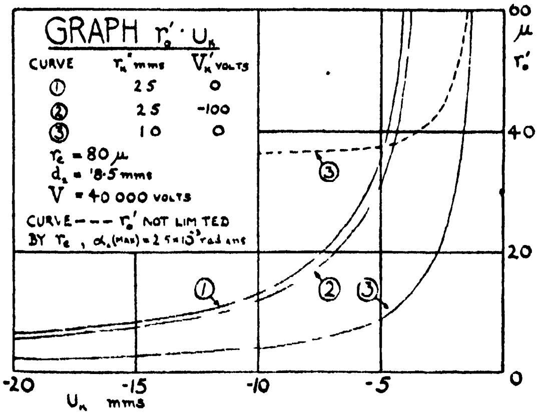


FIG. 9.

trajectories of electrons which leave the filament with zero axial velocity have been considered above. The extension of the theory to cover those electrons which have axial velocities $V_T$ (in electron volts) at the filament would not be difficult but is not attempted here, since the above treatment suffices for the purpose in view. We may note in passing that for the electron gun as commonly employed in the electron microscope $V_T \ll V$, so that the longitudinal chromatic aberration as calculated on the above theory would not be very large.

## THE IDEAL ELECTRON GUN

The heating effect of the beam acting on the specimen is proportional to $\sigma r_0'^2$, where $\sigma$ is the current density at the specimen. The cross-sectional area of film over which this heat can be dissipated is proportional to $r_0'$. There is an obvious advantage in reducing $r_0'$. To a first approximation if $r_0'$ is halved, $\sigma$ may be made four times greater for the same rise in temperature of the specimen. Ideally one would prefer to illuminate only that part of the specimen which is examined on the final screen. If other parts of the specimen are illuminated scattered electrons from these regions may reach the final screen and reduce the contrast.

The illuminating beam, if it does not strike any part of the supporting mesh, will build up a static charge on the film and specimen. Both electron microscopy and electron diffraction require an electron gun giving a small, intense, single source of electrons.

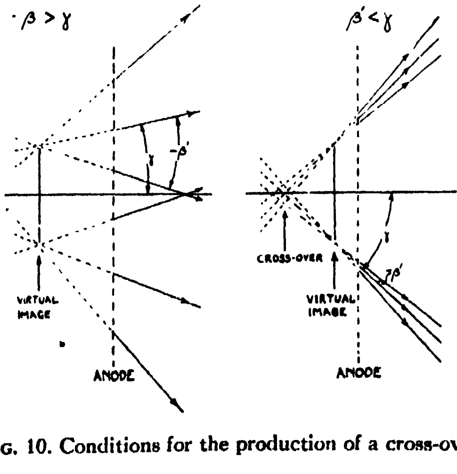


FIG. 10. Conditions for the production of a cross-over.

The angular aperture, $\alpha_c$, of the illumination governs the depth of focus and intensity at constant beam current. When an electron microscope is used for electron diffraction with the specimen below the projector lens the radius of the objective exit pupil which acts as the effective source for the diffraction camera is governed by the aperture of the illumination. The aperture, $\alpha_c$, will influence the resolving power of the microscope, particularly if the objective field lacks symmetry, for in this case when $\alpha_0$ is large the rays leaving the object will move into those parts of the objective field further from the axis with a consequent increase in image assymetries. It has been argued, though convincing experimental proof is lacking, that even with a symmetrical objective field there will be an optimum value for $\alpha_0$ which will be attained for a certain value of $\alpha_c$.

Other factors to be considered in the choice of an electron gun are the ease with which it can be assembled and cleaned. Both the electron gun and condenser lens should form a system which can be readily adjusted to provide an illuminating beam coaxial with the objective lens.

Ideally then one would prefer to have $r_0'$, $\sigma$, and $\alpha_c$ independently variable. In practice, since simplicity of construction is very desirable, this ideal is sacrificed and the chief aim is to provide a small source of adequate intensity.

## PRACTICAL ELECTRON GUNS

The electron guns employed in practice may be divided into two classes according to the conditions under which they are operated. These classes are, the fixed-bias gun operated at emission saturation, and the cathode-biased gun[9]

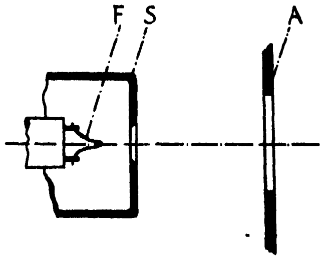


FIG. 11. A three-electrode electron gun as employed in electron microscopy.

---

[9] J. Hillier and R. F. Baker, J. App. Phys. 17, 12 (1946).

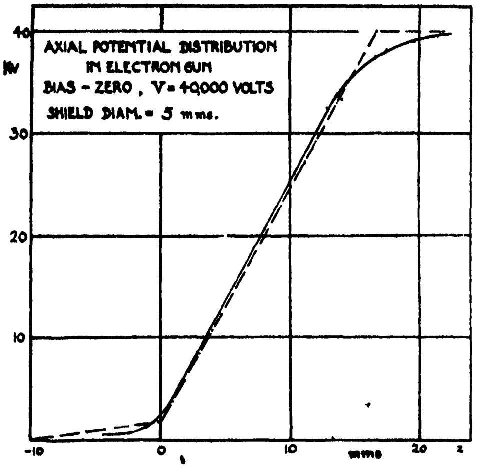


FIG. 12.

operated with space charge limited current.[8] This is not intended to imply that space charge is without effect in the former class of electron gun.

The construction of a fixed-bias gun is shown in Fig. 11. *F* is a hairpin filament of tungsten heated by a d.c. or r-f current. *S* is the shield or grid. It may be maintained at the same potential as the filament or at a negative potential of up to 200 volts with respect to the filament by a battery. *A* is the plate or anode which will be at a voltage of from 30 to 200 kv with respect to the filament in electron microscope practice. Matters of practical interest in the operation of such a gun have been described by Hillier and Baker.[3] Our immediate concern is to see wherein the optical behavior of this type of gun departs from the simple theory outlined above.

The major limitations in the above theory are imposed by the assumption of plane electrodes and linear axial potential variation. With this in mind experiments were undertaken with an electrolytic trough to determine the potential distributions occurring in practice.

Two points may be mentioned in connection with the use of the electrolytic trough. First, since the trough is a relatively large and open structure, it is difficult to shield it from the effect of stray fields which may become troublesome if measurements of high sensitivity are attempted. It is therefore desirable to use for the generator a supply of variable audiofrequency and for the detector a pair of head phones so that the balance can be made to the generator note and stray noises ignored. This is particularly important if an amplifier is used between the probe and the detector, and might well lead to errors if a vacuum-tube voltmeter were employed.

The second point concerns the plotting of the equipotential lines close to the filament. The field distribution near the filament will have a considerable influence on the behavior of the gun since the electrons are moving at slow speeds in this region. It is difficult to determine the equipotentials in this region, however, since the arms of the potential divider have to be made very unequal and the sensitivity of the circuit (which may be considered as a Wheatstone net) becomes very small. Under these conditions the sensitivity was improved by turning a copper plate to the shape of a fairly flat equipotential surface just outside the shield, placing this plate in the tank in the appropriate position, and transfering the lead from the anode to this shaped plate. The low equipotentials could then be studied with less unequal values of the arms of the potential divider.

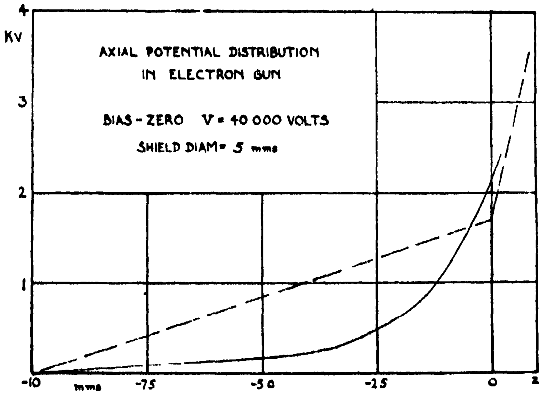


FIG. 13.

Figure 12 shows the axial potential distribution for an electron gun of the form shown in Fig. 11. The full line is from the experimental results; the broken line shows the distribution assumed in the elementary theory above. The origin has been taken at the shield center. In Fig. 13 the distribution of axial potential between the shield and the filament is shown in more detail for $u_k = -10$ mm. Measurements were made with the plane of the filament both in and perpendicular to the surface of the electrolyte. No very significant differences in the axial potential variation were detected in the two cases. It is probable that the magnetic effect of the heating current in the filament would produce effects at least as great as those produced by assymetries in the electric field. For values of $z$ between $-5$ mm and $-10$ mm the measured potentials were of doubtful accuracy, and the curve has been extrapolated linearly to $z=-10$ mm. When $u_k$ was reduced to $-5.0$ mm (Fig. 14) it could be observed that the curve for the axial potential distribution was concave downwards in the vicinity of the filament. In other words, the field strength is relatively high at the filament because of the small radii of curvature of its surface. Under these conditions the optical description of the electron gun is as follows. Leaving the filament the electrons first pass through a thick diverging lens, then a thick converging lens which terminates shortly beyond the shield; at the anode the electron passes through a thick diverging lens. The net effect of these lenses, as compared with the thin lenses treated above will be to further reduce $M_G$, the gun magnification, and to increase the magnitude of $\beta'$, the divergence of the beam from any point of the filament. Under normal conditions of operation, therefore, a cross-over will not be produced in this electron gun.

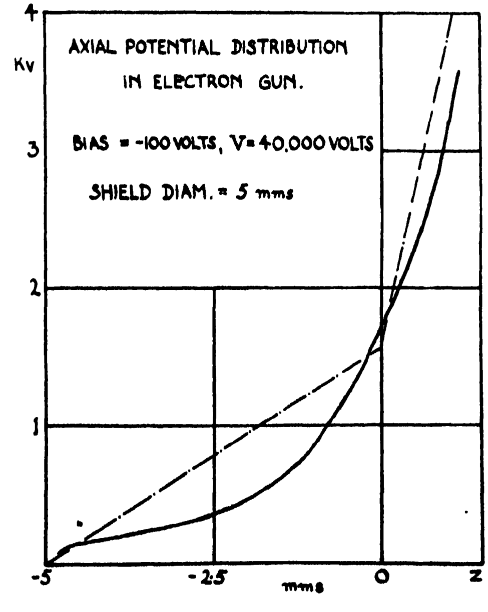


FIG. 14.

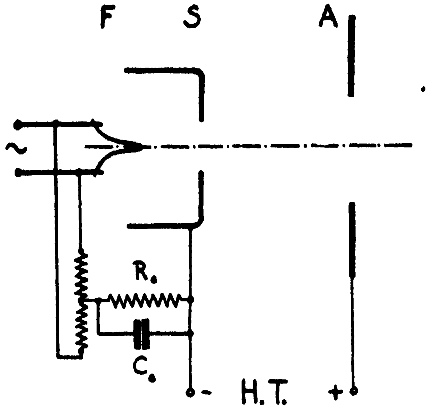


FIG. 15. The cathode-biased electron gun.

Observations with electron microscopes do not discredit this view. In the electron microscope put into operation in 1944 at the University of Toronto[10] the condenser diaphragm can be displaced laterally while the microscope is in operation. If this is done, the region of illumination of the specimen as observed at the intermediate image is seen to move. While this at first sight might seem to indicate the state of affairs shown in Fig. 8, and the existence of a cross-over, it is considered that the effect is due largely to the spherical aberration of the condenser lens.

The major disadvantages of this type of electron gun are the multiple nature of the source, the large source size which must be employed if high intensity is required, and the consequent large beam current. The gun is also inefficient because the large values of $\beta'$ employed mean that most of the electrons hit the condenser diaphragm. The analysis given above indicates a fundamental difficulty in the design of this type of electron gun. The condition for high intensity is $\beta'$ small, which from Eq. 22 indicates

$$\left(\frac{V_T}{V}\right)^{\frac{1}{2}}\cdot\frac{T_0+T_F}{T_F}\quad\text{small.}$$

The condition for a small area of illumination of the specimen is that $M_G$ be small which requires

$$\frac{T_0+T_F}{T_F}\quad\text{to be large.}$$

---

[10] E. F. Burton and W. H. Kohl, *The Electron Microscope* (Reinhold Publishing Corporation, New York, 1946), second edition, Chap. 18.

Unless $V_T/V$ can be reduced or the area of illumination of the specimen independently controlled the final design and operating condition must be a compromise.

### CATHODE-BIASED ELECTRON GUNS

The essential parts of a cathode-biased electron gun are shown in Fig. 15. For small filament and hence small beam currents its optical behavior is similar to that of the fixed-bias gun. As the filament current is increased, however, the beam current reaches a saturation value that is dependent on the geometry of the gun, and the value of the resistor $R_c$. Typical values of the components will be $R_c=0.5$ megohm, $C_c=0.5$ μf, for $r_k''=2.5$ mm. Under these conditions the source is definitely single and intense. Its position is nearly central with respect to the shield and is very little dependent on the centering of the filament with respect to the shield. The image seen at the intermediate screen is no longer the image of the filament but the image of a space charge.

An examination of the image at the intermediate screen of the microscope as the filament current is increased shows the course of events in the electron gun. At low filament current an image of the filament can be seen. It will generally show the die markings on the filament and will be in a position conjugate with the filament. As the filament current is increased this image becomes blurred, larger, and more intense, and moves to a new position, though no movement will be evident if the filament is carefully centered to the axis of the shield and anode holes.

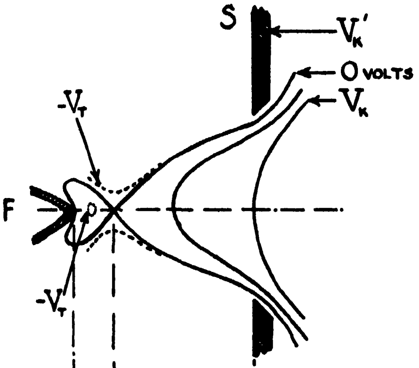


Fig. 16. Assumed potential distribution for a cathode-biased electron gun.

Compared with the zero-biased gun it is possible to obtain a brighter source even when the size of the source and the beam current is smaller. Increasing the shield-to-filament distance reduces the value at which the beam current saturates and also reduces the source size. Over a wide range of filament heights, however, the intensity of illumination is nearly constant, and it is both possible and advantageous to work with the filament raised so that the beam current is of the order 100 to 300 microamperes; the source radius is then of the order 5 to 15 microns and the intensity of the order twenty times that available with the zero-bias gun.

The potential distribution in this type of gun is shown diagrammatically in Fig. 16. Calculations of the potential distribution have been made and an empirical discussion of the behavior of the gun given by Reisner.[11] The space charge enclosed in the zero equipotential surface near the filament acts as a virtual cathode. Since electrons emitted with thermal velocity $V_T$ (in electron volts) cannot cross the $-V_T$ equipotential surface this surface will tend to define the source. Moreover, since electrons cannot reach the shield no secondary sources can be produced by electron reflections there.

The presence of space charge and the ill-defined nature of the virtual cathode do not permit this type of gun to be discussed further in terms of the above theory.

### THE EMISSION MICROSCOPE

Measurements of the performance of an immersion objective have been made by Johannson[12] and Savchenko.[13] In the emission microscope

---

[11] J. H. Reisner, Paper before the Electron Microscope Society of America, December 1946 meeting at Pittsburgh.

[12] H. Johannson, Ann. d. Physik 18, 385 (1933).

[13] F. Savchenko, Zhurnal Tekhnicheskoi Fiziki, Moscow 9, 2211–2219 (1939).

employed by Johannson an image of the emitting cathode was formed on a fluorescent screen 240 mm from the cathode. The constants for the microscope are listed below.

$d_2 = 1.0$ mm
$r_k'' = 0.6$ mm
$v_k = 240$ mm approximately $= L$ a constant for the microscope
$V = 750$ volts.

In operation $u_k$ was set at some value between $-5.0$ mm and zero, and $V_k'$ was adjusted to produce a sharp image on the fluorescent screen. Values of $M_k$, the magnification, and $V_k'$ the screen potential for different values of $u_k$ have been recorded by Johannson.

The foregoing theory must be extended before it can be applied to the above system. Since $v_k \gg d_2$ the electrons travel for the most part at constant speed after leaving the region of the shield. The transit time from the shield to the anode is small and will be ignored compared with the transit time from the anode to the fluorescent screen.

$$T_I = L/\{2d_2^{\frac{1}{2}}(-Ve/2md_2)^{\frac{1}{2}}\}, \qquad (52)$$

and therefore

$$M_k = L[\lambda(1-\epsilon)]^{\frac{1}{2}}/2d_2^{\frac{1}{2}}u_k. \qquad (53)$$

In Fig. 17 the points are calculated from Eq. (53), $\epsilon$ being deduced from the values of $V_k'$ given by Johannson. The agreement between theory and experiment is good for $-u_k > r_k''$. With smaller values of $u_k$ one may expect the assumptions made above as to the potential distribution to become increasingly invalid. In view of the large values of $M_k$, $T_0$ is much smaller than $T_I$, and the condition for focus is therefore approximately

$$T_0 = -T_F. \qquad (54)$$

In calculating the focal time $T_F$ it should be noticed that for this microscope $V_k$ is not always small compared with $V$. Since the electron speed is constant on the image side of the lens, the focal time can be obtained in this case by calculating, first of all, the focal length $F$. If $f_k$, and $f_A$ are the focal lengths of the cathode and anode components, respectively, then:

$$1/F = (1/f_k) + (1/f_A) \qquad (55)$$

$$= \frac{(V-V_k)/d_2 + V_k/u_k}{4V_k} - \frac{(V-V_k)/d_2}{4V}. \qquad (56)$$

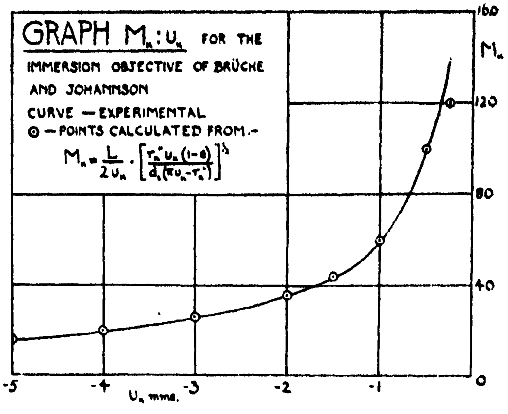


FIG. 17.

Also

$$T_F = F(m/-2eV)^{\frac{1}{2}} \qquad (57)$$

and

$$V_k = r_k'' u_k(1-\epsilon)/[\pi u_k - r_k''(1+u_k/d_2)]. \qquad (58)$$

However Eq. (58) may be replaced by Eq. (13) with sufficient accuracy for the present calculations. With the substitutions previously employed

$$\frac{1}{F} = \frac{[1-\lambda(1-\epsilon)/d_2]^2 u_k + \lambda(1-\epsilon)}{4u_k\lambda(1-\epsilon)}. \qquad (59)$$

Solving for $T_F$ and substituting into Eq. (54)

$$\frac{4\lambda^3(1-\epsilon)^3}{d_2^3} = \left\{\left[1 - \frac{\lambda(1-\epsilon)}{d_2}\right]^2 \frac{u_k}{d_2} + \frac{\lambda(1-\epsilon)}{d_2}\right\}^2. \qquad (60)$$

Working with millimeters as the unit of length we have for the special case of the immersion

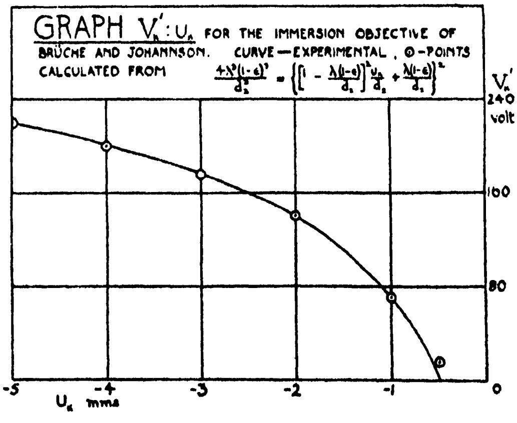


FIG. 18.

*objective used by Johannson $d_2 = 1.0$. Hence writing $x = \lambda(1-\epsilon)$ Eq. (60) becomes*

$$u_k^2x^4 - (4u_k^2 - 2u_k + 4)x^3 + (6u_k^2 - 4u_k + 1)x^2 - (4u_k^2 - 2u_k)x + u_k^2 = 0. \quad (61)$$

Choosing values of $u_k$ Eq. (61) is solved for $x$. From $x$ and the known value of $\lambda$, $\epsilon$ and hence $V_k'$ is deduced. Values of $V_k'$ calculated in this way are plotted as points in Fig. 18—the line represents the values observed by Johannson. The agreement is good for $-u_k > r_k''$.

## CONCLUSIONS

The chief conclusions reached with regard to the zero-bias electron gun used in electron microscopes are that no cross-over is formed in the gun under normal conditions of operation and that the small size of the effective source is due to the demagnification of the gun. The gun is inefficient *since most of the electrons strike the condenser aperture. The dioptrics of such a system is not* Gaussian. The cathode-biased gun is at present the most efficient type of gun for electron microscopy.

A satisfactory first-order theory of the immersion objective has been obtained for the case in which the distance from cathode to shield is greater than the radius of the shield aperture. The dioptrics of this system is also not Gaussian.

## ACKNOWLEDGMENTS

It is a pleasure to record my indebtedness to Professor E. F. Burton, Director of the McLennan Laboratory for his encouragement during the prosecution of this work. Mr. R. S. Sennett, also of the McLennan Laboratory gave valuable help by making the electrolytic trough determinations. Finally I wish to thank Dr. James Hillier of the RCA Laboratories for useful information and informative discussions on the cathode-biased gun.

# On the One-Dimensional Theory of Steady Compressible Fluid Flow in Ducts with Friction and Heat Addition*

BRUCE L. HICKS
*Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland*
DONALD J. MONTGOMERY
*Office of Naval Research (London Branch)*

AND

ROBERT H. WASSERMAN
*University of Chicago, Chicago, Illinois*

(Received April 17, 1947)

Steady, diabatic (non-adiabatic), frictional, variable-area flow of a compressible fluid is treated in differential form on the basis of the one-dimensional approximation. The basic equations are first stated in terms of pressure, temperature, density, and velocity of the fluid. Considerable simplification and unification of the equations is then achieved by choosing the square of the local Mach number as one of the variables to describe the flow.

The transformed system of equations thus obtained is first examined with regard to the existence of a solution. It is shown that, in general, a solution exists whose calculation requires knowledge only of the variation with position of any three of the dependent variables of the system. The direction of change of the flow variables can be obtained directly from the transformed equations without integration. As examples of this application of the equations, the direction of change of the flow variables is determined for two special flows.

In the particular case when the local Mach number $M=1$, a special condition must be satisfied by the flow if a solution is to exist. This condition restricts the joint rate of variation of heating, friction, and area at $M=1$. Further analysis indicates that when a solution exists at this point it is not necessarily unique.

Finally, it is shown that the physical phenomenon of choking, which is known to occur in certain simple flow situations, is related to restrictions imposed on the variables by the form of the transformed equations. The phenomenon of choking is thus given a more general significance in that the transformed equations apply to a more general type of flow than has hitherto been treated.

## INTRODUCTION

THE rational experimental development of jet- and rocket-propulsion power plants requires adequate knowledge of the theoretical mechanics of diabatic (non-adiabatic), frictional, variable-area compressible fluid flow. The differential equations describing this type of flow are well known.[1a, 1b, 2–4] Their solution in the three-dimensional case, however, is so difficult that some simplification is necessary to permit development of the theory in a form immediately useful for technical applications.

In the present paper, such simplification is effected by generalizing the familiar "one-dimensional" or hydraulic treatment of steady fluid flow to include the simultaneous effects of heat addition, friction, and area change upon the flow of a compressible fluid rather than by attempting to show that the one-dimensional approximation follows from a simplification of the hydrodynamic and heat-flow equations in their general three-dimensional form. The generalization leads to one-dimensional equations in differential form, which are identical with equations previously used by other investigators in less general cases.

Generalized conservation equations have been derived in Appendix A in order that a complete and logical basis for the theory may be accessible to the reader. The resulting theory is intended to serve as a foundation in differential form for calculation of all types of mathematically continuous (that is, shockless) flow of perfect gases to which the one-dimensional approximation is applicable. Thus the theory applies directly to

---

* This paper is a revised report of theoretical work performed by the authors at the Cleveland Laboratory of the National Advisory Committee for Aeronautics in 1944–45, and was issued in slightly different form as NACA TN No. 1336, 1947.

[1] W. F. Durand, ed. *Aerodynamic Theory* (Julius Springer Verlag, Berlin, 1935): (a) Vol. I, p. 237; (b) Vol. III, pp. 47–49.

[2] C. Eckart, Phys. Rev. **58**, 267 (1940).

[3] Wien-Harms, *Handbuch der Experimental Physik* (Akaddemische Verlagsgesellschaft, Leipzig, 1931), Bd. 4, pp. 343–369.

[4] A. Vazsonyi, Quart. Appl. Math. **3**, 29 (1945).

compressible flow in combustion chambers and also, with but slight modification, to flow in turbines, compressors,[5] nozzles, and diffusers, whenever the one-dimensional approximation is valid.

In order to obtain convenient and unified equations, the generalized relations are transformed by introducing a new basic variable, the square of the local Mach number $M^2=N$. Pressure and temperature are chosen as the additional basic variables; other relevant flow variables (for example, density, velocity, mass flow) may be expressed in terms of Mach number squared, pressure, and temperature. Values of $M$ from zero to infinity are considered; the treatment is therefore applicable to both subsonic and supersonic flow.

The variable $M$ has been used throughout differential treatments by Gukhman,[6] Bailey,[7] Nielsen,[8] and Chambre and Lin,[9] who investigated various examples of frictional, diabatic compressible flow.** A related variable

$$Z=(\gamma-1)M^2/[2+(\gamma-1)M^2],$$

which can be used alternatively with $M$, will be discussed briefly in Appendix B. Pertinent papers in which $M$ is not used extensively are those by Binder[10] and Keenan and Neumann,[11] which report studies of isothermal and adiabatic frictional flow, respectively. A treatment of frictionless, diabatic compressible flow carried out by Szczeniowski[12] is partly in differential form. The same subject, in $M$ language without differential formulation, is discussed by Hicks[13] and Wood.[14] The variable $M$ has also been employed to advantage by Sorg[5] for the analysis of compressible flow through turbines and compressors, a related field that is not specifically discussed in the present paper. Mach and Crocco vectors, the square of whose magnitudes are $N$ and $Z$, have been found useful for the description of three-dimensional adiabatic and diabatic flows.[15a, 15b]

In the general case, the differential equations obtained in the present treatment do not permit of formal integration; but being of first order, they are particularly amenable to numerical methods. A solution of the system is shown to exist, except possibly at sonic velocity, and the behavior of the solution in this neighborhood is investigated. From the differential equations useful information may easily be obtained about direction of changes in the flow variables. Choking is shown to be a consequence of a certain property of the equations.

## THE ONE-DIMENSIONAL APPROXIMATION

### Basic Equations

The "one-dimensional" steady-flow theory utilizes a model consisting of a perfect gas contained within a duct, across any section of which the flow variables are constant. Only the component of velocity normal to the section is considered; body forces are neglected, and heat, whether supplied by combustion, conversion of frictional work, or conduction from the walls, is assumed to be transferred instantaneously and completely but only transversely throughout the cross section, which may be of variable area. Each flow variable can thus be considered as a function of a single parameter, say the distance along the axis of the tube, whence the term "one-dimensional."

The conventional variables—pressure, temperature, density, and velocity in one-dimensional flow—are connected by four relations derivable from the first law of thermodynamics, the conservation of mass, the second law of motion, and the thermal equation of state for a perfect gas. The four relations are:

---

[5] K. W. Sorg, Forschung 10, 270 (1939).

[6] A. A. Gukhman, J. Tech. Phys. U.S.S.R. 9, 411 (1939).

[7] N. P. Bailey, J. Ae. Sci. 2, 227 (1944).

[8] J. N. Nielsen, NACA ARR No. L4C16 (1944).

[9] P. Chambre and C. C. Lin, J. Ae. Sci. 13, 537 (1946).

** The authors regret that explicit reference could not be made to several valuable (classified) papers by E. R. Hawthorne. It is our understanding that Dr. Hawthorne and Professor Shapiro have been invited to present a comprehensive paper on flow of compressible fluids with heat transfer at a meeting of the American Society of Mechanical Engineers later this year.

[10] R. C. Binder, A.S.M.E. Trans. 66, 221 (1944).

[11] J. H. Keenan and E. P. Neumann, NACA Tech. Note No. 963 (1945).

[12] B. Szczeniowski, Can. J. Research 23, 1 (1945).

[13] B. L. Hicks, NACA ACR No. E5A29 (1945) (Wartime Report No. E-88).

[14] G. P. Wood, J. Ae. Sci. 14, 24 and 63 (1947).

[15] (a) B. L. Hicks, P. E. Guenther, and R. H. Wasserman, Quart. Appl. Math. (1947); (b) B. L. Hicks, Quart. Appl. Math., accepted for publication in 1948, see also Phys. Rev. 69, 135 (A), 250 (A) (1946); 71, 476 (A) (1947); Ballistic Research Laboratories Report No. 633 (1947).

Conservation of energy $c_p dT + VdV = dQ$, (1)

Conservation of mass $d(\rho VA) = 0$, (2)

Equation of motion $-dp = \rho VdV + \rho dF$, (3)

Equation of state $d(p/R\rho T) = 0$. (4)

The specific heat at constant pressure $c_p$ and the gas constant $R$ do not vary in the flow. The symbols $\rho$, $V$, $p$, and $T$, respectively, stand for density, velocity, absolute static pressure, and absolute static temperature. The pipe area, which may be variable, is represented by $A$. Heat added per unit mass is indicated by $Q$, and work per unit mass done against friction by $F$. Consistent units are used throughout. In Eqs. (1)–(4) each variable is to be considered as a function of a single parameter, such as the distance $x$ along the tube considered positive in the direction of flow; and, of course, the meaning of each differential $du$ is then given by

$$du = u'(x)dx.$$

Equations (1)–(3) are customarily used without explicit recognition of their true meaning with regard to the one-dimensional approximation. The interpretation of the quantity $dF$ in particular is often obscure. In order to provide a logical, unified basis for the theory, Eqs. (1)–(3) are derived in Appendix A; special care is taken to keep the derivations within the framework of the one-dimensional approximation.

### Applicability

The validity of the one-dimensional approximation depends upon the assumption of the uniformity of flow conditions across a plane normal to the direction of flow. Experience has shown that this assumption constitutes an adequate approximation in many special cases; in particular, with subsonic turbulent flow in long pipes without separation, the reasonably flat velocity profile permits the use of equations derived on this basis. Van Driest[16] has shown how the results of one-dimensional theory for incompressible fluids can be corrected for the effects of turbulence and non-uniformity of velocity distribution in cases where these factors have been evaluated experimentally. The corrections to the energy equation are quite complicated, and their complexity would be increased for compressible flow. In general there is insufficient experimental data available at present to permit formulation of such corrections where compressibility, incomplete growth of the boundary layer, or separation of the flow must be considered. Although compressibility and boundary layer effects are somewhat amenable to calculation, the occurrence of sparation is difficult or impossible to predict, and the question of applicability, in these cases, of the one-dimensional approximation must be determined by experiment or estimated by experience.

The one-dimensional approximation would not be valid if oblique shocks occur in the flow. Nor can normal shocks, if treated as flow discontinuities, be handled in the differential form of the present approximation. If, however, in Eqs. (1) and (3), $dQ$ and $dF$ are considered to depend upon the derivatives of $T$ and $V$ and if heat and momentum transfer in the direction of flow is allowed, then the equations for continuous normal shock[17] can be put in the form of Eqs. (1)–(4).

In the development and use of Eqs. (1)–(4) various approximations are made, such as neglecting the squares of velocity components normal to the direction of flow, replacing the square of the cosine of the half-angle by unity, and assuming the constancy of $R$ and $c_p$. In this paper no attempt is made to state under what circumstances such approximations are suitable.

## TRANSFORMATION OF EQUATIONS

### Change of Variables

A canonical form for Eqs. (1)–(4) is obtained by taking logarithmic derivatives and choosing as a variable the square of the local Mach number,

$$M^2 \equiv N = V^2/\gamma RT, \qquad (5)$$

where $\gamma$ is the ratio of the specific heats. This choice to obtain simplification of the equations is not unique; similar advantages result with other dimensionless combinations of velocity squared and a temperature. For instance, some workers have used the ratio of dynamic tem-

---

[16] E. R. Van Driest, A.S.M.E. Trans. **68A**, 231 (1946).

[17] Reference 1, p. 219.

perature to total temperature; in Appendix B to the present report, the canonical differential equations in terms of this variable are given.

If Eqs. (1)–(4) are divided by $c_pT$, $\rho VA$, $p$, and $p/R\rho T$, respectively, there result

$$(\gamma-1)\frac{V^2}{\gamma RT}\frac{dV}{V}+\frac{dT}{T}=\frac{dQ}{c_pT}, \tag{6}$$

$$\frac{dV}{V}+\frac{d\rho}{\rho}=-\frac{dA}{A}, \tag{7}$$

$$\gamma\frac{V^2}{\gamma RT}\frac{dV}{V}+\frac{dp}{p}=-\frac{\rho dF}{p}, \tag{8}$$

$$-\frac{dp}{p}+\frac{dT}{T}+\frac{d\rho}{\rho}=0. \tag{9}$$

With use of Eq. (5) and the expression for $dV/V$ obtained by logarithmic differentiation of Eq. (5),

$$\frac{dV}{V}=\frac{1}{2}\left(\frac{dN}{N}+\frac{dT}{T}\right) \tag{10}$$

and, upon elimination of $d\rho/\rho$, there are found

$$-\frac{(\gamma-1)N}{2}\frac{dN}{N}+\left[1+\frac{(\gamma-1)N}{2}\right]\frac{dT}{T}=\frac{dQ}{c_pT}\equiv d\theta, \tag{11}$$

$$\frac{1}{2}\frac{dN}{N}+\frac{dp}{p}-\frac{1}{2}\frac{dT}{T}=$$

$$-\frac{dA}{A}\equiv d\alpha, \tag{12}$$

$$-\frac{\gamma N}{2}\frac{dN}{N}-\frac{dp}{p}-\frac{\gamma N}{2}\frac{dT}{T}=\frac{dF}{RT}\equiv d\mu, \tag{13}$$

where the dimensionless quantities $d\theta$, $d\alpha$, and $d\mu$ have been introduced to simplify the following analysis.

### Solution for Logarithmic Differentials

If the determinant formed by the coefficients of $dN/N$, $dp/p$, and $dT/T$ in Eqs. (11), (12), and (13) is not identically zero, the equations may be solved uniquely for these three differentials. As the determinant in question is proportional to $(1-N)$, which vanishes only for $N=1$, the solution is obtained as follows:

$$\begin{aligned}dN/N=(1-N)^{-1}\{(1+\gamma N)d\theta+[2+(\gamma-1)N]d\mu\\+[2+(\gamma-1)N]d\alpha\},\end{aligned} \tag{14}$$

$$\begin{aligned}dp/p=(1-N)^{-1}\{-\gamma Nd\theta\\-[1+(\gamma-1)N]d\mu-\gamma Nd\alpha\},\end{aligned} \tag{15}$$

$$\begin{aligned}dT/T=(1-N)^{-1}[(1-\gamma N)d\theta-(\gamma-1)Nd\mu\\-(\gamma-1)Nd\alpha].\end{aligned} \tag{16}$$

It is also convenient to record the differential expressions for the density $\rho$ and velocity $V$:

$$\begin{aligned}d\rho/\rho=dp/p-dT/T=(1-N)^{-1}\\\times(-d\theta-d\mu-Nd\alpha),\end{aligned} \tag{17}$$

$$\begin{aligned}dV/V=(dN/N+dT/T)/2\\=(1-N)^{-1}(d\theta+d\mu+d\alpha).\end{aligned} \tag{18}$$

### Application of Second Law of Thermodynamics

The first law of thermodynamics was used in the formulation of the basic equations; the second law of thermodynamics may be employed to furnish additional information. The entropy differential $dS$ for a perfect gas is given[18] by

$$\begin{aligned}dS/c_p=dT/T-[(\gamma-1)/\gamma]dp/p\\=d\theta+[(\gamma-1)/\gamma]d\mu.\end{aligned} \tag{19}$$

The second law of thermodynamics then states

$$0\leqslant dS/c_p-dQ/c_pT=[(\gamma-1)/\gamma]d\mu. \tag{20}$$

The relation, according to Eq. (19), that

$$dS/c_p=(d\theta+d\mu+d\alpha)-d\mu/\gamma-d\alpha$$

when used with Eq. (20), results in the inequalities

$$d\theta\leqslant dS/c_p=d\beta-d\mu/\gamma-d\alpha\leqslant d\beta-d\alpha, \tag{21}$$

where $d\beta\equiv d\theta+d\mu+d\alpha$.

## DISCUSSION OF EQUATIONS

### Remarks on Integration of Equations

Equations (14)–(16) can be rewritten as

$$\begin{aligned}N'=\frac{N}{1-N}\frac{1+\gamma N}{c_pT}Q'+\frac{N}{1-N}\frac{2+(\gamma-1)N}{RT}F'\\-\frac{N}{1-N}\frac{2+(\gamma-1)N}{A}A',\end{aligned} \tag{22}$$

---

[18] P. S. Epstein, *Textbook of Thermodynamics* (John Wiley and Sons, Inc., New York, 1937), p. 63.

$$p' = -\frac{p}{1-N}\frac{\gamma N}{c_pT}Q' - \frac{p}{1-N}\frac{1+(\gamma-1)N}{RT}F' + \frac{p}{1-N}\frac{\gamma N}{A}A', \quad (23)$$

$$T' = \frac{T}{1-N}\frac{1-\gamma N}{c_pT}Q' - \frac{T}{1-N}\frac{(\gamma-1)N}{RT}F' + \frac{T}{1-N}\frac{(\gamma-1)N}{A}A', \quad (24)$$

where the primes indicate differentiation with respect to $x$. This system clearly satisfies, except at $N=1$, the conditions of the fundamental existence theorem[19] when $Q$, $F$, and $A$ are differentiable. Hence a solution exists except at sonic velocity and may be obtained formally when possible, and by standard numerical methods otherwise, as soon as the functions $Q$, $F$, and $A$ (or their derivatives) are specified. More generally, the system may be solved in similar fashion for any three of the variables $N$, $p$, $T$, $Q$, $F$, and $A$ as functions of $x$, when the variation with $x$ of the other three is prescribed. Also it may be noted that as all the foregoing variables are functions of one parameter, any two may be considered as functions of each other under suitable circumstances.

### Direction of Change of Flow Variables

In practical as well as in theoretical work it is frequently useful to be able to determine the direction of change of flow quantities with respect to heat addition, friction, or area variation without troubling to get quantitative information from integrated forms. Equations (14)–(18) or (22)–(24) permit the specification of signs of derivatives at any particular point and also throughout certain regions of flow. Thus Eq. (14) shows that in subsonic flow the effect of positive $\theta'$, $\mu'$, or $\alpha'$, is to increase $N$, whereas for supersonic flow the effect is to decrease $N$. When the derivatives have different signs, the net effect will depend upon the algebraic sum of the separate contributions.

As an example of the use of this technique, suppose heat is added to a fluid in a constant-area pipe, with negligible friction; that is, $\theta' \neq 0$, $\mu' = \alpha' \equiv 0$. It is easily seen from Eqs. (14) through (18) that for the entire range of $N$ from zero to infinity

$$(1-N)dN/dQ \geqq 0, \quad (25)$$

$$dp/dN \leqq 0, \quad (26)$$

$$(1-\gamma N)^{-1}dT/dN \geqq 0, \quad (27)$$

$$d\rho/dN \leqq 0, \quad (28)$$

$$dV/dN \geqq 0. \quad (29)$$

These results are given by Hicks.[13] By use of the chain rule for the derivative of a function of a function, the sign of the derivative of any of the flow variables with respect to any of the others may be obtained; thus, from Eqs. (25) and (29) it is clear that

$$(1-N)\frac{dV}{dQ} = (1-N)\frac{dV}{dN}\frac{dN}{dQ} \geqq 0. \quad (30)$$

As another example, consider the flow in circular cylindrical pipes with heat addition and with friction; that is, $\theta' \neq 0$, $\mu' \neq 0$, $\alpha' \equiv 0$. (See also related discussion by Nielsen.)[8] Equation (14) will be used to determine the direction of change of $N$ with respect to $x$. If the heat addition is only through the wall, which is at temperature $T_w$, the heat added per unit mass of fluid in passing a distance $dx$ along the tube is given by

$$\rho V A dQ = h(T_w - T)(\pi D dx), \quad (31)$$

where $D$ is the tube diameter, and $h$ the local surface-to-fluid coefficient of heat transfer in heat units transferred per unit temperature difference, per unit area.[20a] In conjunction with Eq. (11), Eq. (31) leads to

$$d\theta \equiv \frac{dQ}{c_pT} = \frac{h[(T_w/T)-1]\pi D dx}{c_p\rho V\pi D^2/4} = \frac{4h[(T_w/T)-1]dx}{c_p\rho VD}. \quad (32)$$

The expression for frictional work done is

---

[19] L. Bieberbach, *Theorie der Differentialgleichungen* (Julius Springer Verlag, Berlin, 1930), pp. 34–35.

[20] W. H. McAdams, *Heat Transmission* (McGraw-Hill Book Company, Inc., New York, 1942). (a) Eq. 2, p. 135; (b) Eq. 8, p. 119; (c) Eq. 1, p. 162.

assumed to be given[20b] (see also Appendix A) by

$$dF=\frac{fV^2dx}{2(D/4)},\tag{33}$$

where $f$ is the Fanning friction factor. From Eqs. (13) and (33) it follows that

$$d\mu\equiv\frac{dF}{RT}=\frac{2f\gamma Ndx}{D}.\tag{34}$$

If Reynolds' analogy is valid, $h$ may be replaced[20c] by $c_p\rho Vf/2$, whence Eq. (14) becomes

$$dN/dx=[N/(1-N)]\{(1+\gamma N)[(T_w/T)-1] +[2+(\gamma-1)N]\gamma N\}2f/D.\tag{35}$$

This equation may be used to determine the direction of change of $N$ with $x$, and hence of other flow quantities, for various ranges of $N$ and of $T_w/T$. For values of $(T_w/T)\ll 1$, (maximum rate of cooling), $dN/dx$ is positive for values of

$$1>N>[-\gamma+(\gamma(5\gamma-4))^{\frac{1}{2}}]/2\gamma(\gamma-1)=0.58$$

for $\gamma=1.4$; that is, the effects of friction in increasing the Mach number overbalance the effects of the cold walls in lowering it if $1>M=N^{\frac{1}{2}}>0.76$ for $\gamma=1.4$ If $N>1$ then $(dN/dx)<0$, and acceleration of frictional, supersonic flow by convective cooling appears to be impossible. This behavior of compressible gases affects the design of supersonic wind tunnels.[21] Acceleration of frictionless supersonic flow by cooling should, however, be possible[9,13].

## Behavior of Solution at Sonic Velocity

The differential equations (14)–(18) must be examined as to behavior at the singular point $N=1$. In order that the logarithmic differentials may be defined at this point, it is necessary that

$$d\beta\equiv d\theta+d\mu+d\alpha=0 \text{ at } N=1,\tag{36}$$

because each logarithmic differential is proportional to $d\beta/(1-N)$ there. If $d\beta\neq 0$ upstream of the end of the duct, $N$ can equal 1 only at the end of the duct. This situation is illustrated by the "choked" converging nozzle and by the frictional diabatic flow, which is treated in the previous section. Equation (36) is formally satisfied at the end of a duct where $d\theta$, $d\mu$, and $d\alpha$ may be considered to vanish for all values of $N$.

Between the ends of a duct, however, $d\beta$ must always vanish where $N=1$. This condition shows that at $N=1$ arbitrary variations of $d\theta$, $d\mu$, and $d\alpha$ are not possible. A specific illustration is the ideal nozzle in which $d\theta=d\mu=0$; according to Eq. (36), $d\alpha$ is then restricted to the value 0, which means that the area has a stationary value at $N=1$. This is the well-known result that sonic velocity can be attained only in the throat of an ideal nozzle in shockless flow. A quite similar treatment applies for the cases where $d\theta$ and $d\mu$ are the quantities to be investigated (see pertinent material in references 5–14 and 21). Condition (36), which was necessitated by the presence of the determinant $(1-N)/2$ in Eqs. (14)–(18), is thus seen to provide a unification of the treatment of the flow behavior in the neighborhood of sonic velocity.

Combination of the second law of thermodynamics with Eq. (36) also yields limitations on the behavior of the flow at $N=1$. According to Eqs. (21) and (36), at sonic velocity

$$d\theta\leqq dS/c_p\leqq -d\alpha=dA/A.\tag{37}$$

These results may be stated in words to the effect that in converging or constant-area channels at $N=1$, neither the heat term $d\theta=dQ/c_pT$ nor the entropy term $dS/c_p$ can be positive. In diverging channels these two terms may be either positive or negative. If either $d\theta$ or $d\mu$ is everywhere 0, relation (21) yields more detailed results. For example, if $d\theta\equiv 0$, then at $N=1$, by Eqs. (36) and (20)

$$\begin{aligned}d\mu&=-d\alpha\\&=[\gamma/(\gamma-1)]dS/c_p\geqq 0,\quad (d\theta=0,\ N=1).\end{aligned}$$

Continuous flow with friction and without heat addition at sonic velocity cannot, therefore, occur in a converging channel.[7]

A more complete treatment of the behavior of the flow when $N=1$ between the ends of the duct is obtained by considering second-order terms. As $N$ approaches unity, Eq. (14), which may be written

$$N'/N=\{(1+\gamma N)\theta'+[2+(\gamma-1)N]\mu' +[2+(\gamma-1)N]\alpha'\}/(1-N),$$

---

[21] F. Clauser, Phys. Rev. 71, 465A (1947).

takes the form 0/0. For the evaluation of this limit, L'Hopital's rule gives after some calculation

$$N_0'=[\theta_0'N_0'+(1+\gamma)\beta_0'']/(-N_0'),$$

where subscript 0 denotes value of function at $N=1$. The solution for $N_0'$ is

$$N_0'=-\theta_0'/2\pm[(\theta'_0/2)^2-(1+\gamma)\beta_0'']^{\frac{1}{2}}. \quad (38)$$

The double-valuedness of the derivative at $N=1$ will have important consequences in that a unique solution of the equations may not be obtained when $N=1$ along the flow path. In general, it will be possible to continue the solution from $N=1$ along either of two paths, depending on the choice of sign. In certain cases, depending on the signs of $\theta_0'$ and $\beta_0''$, one sign will correspond to continuation into subsonic flow, the other into supersonic; otherwise the two choices will correspond to different continuations into flow of the same character. This result means that specification of initial conditions and of variation of $d\theta$, $d\mu$, and $d\alpha$ alone is not sufficient to insure a unique solution if $N$ becomes unity along the flow. In the event that the radicand is zero, it is possible that only one solution is obtained; or it may happen that some higher derivative is double-valued with resulting ambiguity of solution. The analysis for this case is somewhat involved and will not be continued here.

It is interesting to note that a less general problem of the same nature has been presented by Lorenz[22] and Prandtl and Proell.[23] Some of this work is possibly more accessible in Stodola's book.[24]

## The Phenomenon of Choking

The general equations (11)–(13) impose restrictions on the relations between the flow variables and the heat, friction, and area variation. When these restrictions take the form of upper or lower limits on mass flow, the associated phenomena are termed "choking" processes. As an example, it is well known that the ideal nozzle has for given subsonic entry conditions a maximum mass flow beyond which the discharge cannot be increased no matter how much the exit pressure is lowered. Another case is "thermal choking," wherein the entrance Mach number and hence mass flow in diabatic, frictionless, constant-area flow is limited for given heat addition despite indefinite reductions in outlet pressure.[13]

The nature of choking may be studied with the help of Eq. (22), which was derived simply from the basic equations. It will be shown that unless heat, friction, and area variation are such that $(1-N)$ times the right-hand side of Eq. (22) changes from positive to negative as $x$ increases, the Mach number in the tube cannot become greater than 1 if the entrance velocity is subsonic and cannot become less than 1 if the entrance velocity is supersonic, provided that the flow variables remain continuous.

For convenience, designate by $Y$ the factor $(1-N)$ times the right-hand side of Eq. (22). The quantity $Y$ is seen to consist of a sum of terms in $Q'$, $F'$, and $-A'$ multiplied by functions of $N$ that are always positive. (In the event that only one of the terms $Q'$, $F'$, and $-A'$ is not 0, $Y$ becomes merely the derivative of the heat added, the frictional work, or the area, multiplied by a simple function of the flow variables; then positive $Y$ corresponds to the case of heat addition, friction, or a converging duct.)

Suppose first that $Y$ is always negative. Then if the flow at the entrance section $x_1$ is subsonic, $dN/dx=(1-N)^{-1}Y<0$, and the Mach number decreases; if the entering flow is supersonic, $dN/dx=(1-N)^{-1}Y>0$, and the Mach number increases.

Suppose now that $Y$ is always positive. Then, if the entering flow at $x_1$ is subsonic, $dN/dx=(1-N)^{-1}Y>0$, and the Mach number increases. But $N$ cannot increase past unity as $x$ increases. For suppose $N=1$ at $x=x_0$ and is greater than 1 in the right-hand neighborhood of $x_0$ (exclusive of $x_0$); then $dN/dx$ is negative in this neighborhood, because $(1-N)$ is less than 0 and $Y$ is greater than 0. Now $N$ is equal to unity at $x=x_0$, is continuous, and has a negative derivative in the neighborhood mentioned. Hence $N$ is less than 1 in this neighborhood, which contradicts the assumption. Therefore $N$ cannot be

---

[22] H. Lorenz, Physik. Zeits. 4, 333 (1903).

[23] L. Prandtl and R. Proell, V.D.I. Zeits. 48, 348 (1904).

[24] A. Stodola, *Steam and Gas Turbines* (McGraw-Hill Book Company, Inc., New York, 1927), pp. 98–101.

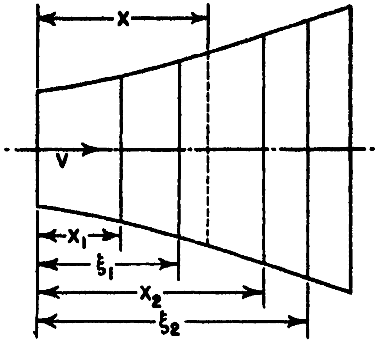


FIG. 1. Fluid elements considered in deriving the equations of energy and mass conservation.

greater than 1 if $Y$ is always positive, $N$ is continuous, and $N(x_1)$ is less than 1. In general, no continuous solution exists for values of $x > x_0$ if $Y$ is always positive. This statement, and the foregoing proof, are valid even if $N'(x_0)$ does not exist. An analogous development may be made for $N(x_1)$ greater than 1 with the conclusion that, with $Y$ positive and $N$ continuous, $N$ cannot be less than 1.

If $Y$ changes from positive to negative at $x = x_0$, however, the value of $N$ may cross unity at that point, but if $Y$ is initially negative, $N$ goes away from unity as previously shown and can only turn toward unity if $Y$ changes from negative to positive; this change must be made at some value of $N$ other than 1. After $Y$ has changed to positive, the situation reduces to the case that $Y$ is always positive, with the $N$ where $Y$ changes sign now taken as the entrance $N$.

It has been shown that up to some fixed point in the tube, which can be either the exit or the point at which $Y$ changes from positive to negative, $N$ and, therefore, the Mach number do not become greater than 1 if the entering velocity is subsonic nor less than 1 if the entering velocity is supersonic. Furthermore, if $Y$ is positive up to the point at which $N$ is limited, the derivative of $N$ before this point is always positive if the entering flow is subsonic and is always negative if the entering flow is supersonic. Thus, for positive $Y$ and subsonic entrance velocity the entrance $N$ cannot exceed some limiting value less than 1 determined by the particular $Q$, $F$, $-A$ variation, for $N$ is always increasing from its initial value and cannot exceed 1; and, by analogous considerations for positive $Y$ and supersonic entrance velocity $N$ cannot be less than some limiting value greater than 1. This limitation is essentially the choking phenomenon.

The specific form this limitation takes is not easily stated in the general case, because the choice of which factors are to be held constant and of which variables are to be considered limited determines the particular form of the restrictions. Numerical results for some particular cases are given in references 7, 13, and 14, among others, to illustrate the nature of possible results. It is felt that additional special cases should be investigated before a thorough study of the general case is attempted.

## APPENDIX A

### Derivations of Three Basic Equations

#### *Conservation of Energy*

The first law of thermodynamics applies to energy changes between two states of a system enclosed within a surface. Let the system (Fig. 1) be the gas of mass $\Delta m$ that is contained in the initial state within the tube walls and the sections at $x_1$ and $x_2$ and in the final state within the tube walls and the sections at $\xi_1$ and $\xi_2$; $x_1$, $x_2$, and $\xi_1$ are arbitrary, and $\xi_2$ is determined by the condition that the mass between $x_1$ and $\xi_1$ equal the mass between $x_2$ and $\xi_2$. When $m(x)$ is defined as the total mass of gas contained within the duct between the sections $x=0$ and $x=x$, the definition of $\Delta m$ becomes

$$\Delta m \equiv m(x_2) - m(x_1) = m(\xi_2) - m(\xi_1). \quad (39)$$

Let $U(x)$ denote the internal (thermal) energy per unit mass of fluid, $p(x)$ the static pressure, $V(x)$ the gas velocity, $A(x)$ the cross-sectional area of the duct, each taken at the section $x$; and $Q(x)$ the heat (in mechanical-energy units) added from walls or by combustion*** to a unit mass of the fluid during its passage from $x=0$ to $x=x$. Then the first law of thermodynamics says for

---

*** Actually, the heat liberated by combustion might be considered as part of the internal energy; or the external surroundings might be considered to include the fuel; or the first law might be generalized to include heat sources. The treatment given here is convenient but must be understood to require some justification on one of the bases mentioned.

steady flow that

$$\int_{m(\xi_1)}^{m(\xi_2)} (U+\tfrac{1}{2}V^2)dm - \int_{m(x_1)}^{m(x_2)} (U+\tfrac{1}{2}V^2)dm$$
$$= \int_{m(\xi_1)}^{m(\xi_2)} Qdm - \int_{m(x_1)}^{m(x_2)} Qdm$$
$$-\left[\int_{x_2}^{\xi_2} pAdx - \int_{x_1}^{\xi_1} pAdx\right]. \quad (40)$$

As $Adx = (1/\rho)dm$, where $\rho(x)$ is the density of the gas, Eq. (40) becomes

$$\int_{m(x_2)}^{m(\xi_2)} (U+\tfrac{1}{2}V^2-Q+p/\rho)dm$$
$$= \int_{m(x_1)}^{m(\xi_1)} (U+\tfrac{1}{2}V^2-Q+p/\rho)dm \quad (41)$$

when the limits of integration are changed.

Provided that the integrand is continuous (which requirement excludes shock), the expression on the right-hand side of Eq. (41) may be written, by the theorem of the mean for integrals:

$$[m(\xi_1)-m(x_1)]f(m_1^*),$$

where $f(m_1^*)$ indicates the value of the integrand at some $m=m_1^*$, $m(x_1)<m_1^*<m(\xi_1)$. The integral on the left-hand side yields a similar result, with subscript 1 replaced by subscript 2, whence, by virtue of Eq. (39)

$$f(m_1^*)=f(m_2^*); \quad m(x_1)<m_1^*<m(\xi_1),$$
$$m(x_2)<m_2^*<m(\xi_2). \quad (42)$$

As $\xi_1 \rightarrow x_1$, $\xi_2 \rightarrow x_2$, this equation becomes

$$f[m(x_1)]=f[m(x_2)]. \quad (43)$$

But $x_1$ and $x_2$ are arbitrary. Hence $f$ is a constant, and

$$\frac{d}{dx}(U+\tfrac{1}{2}V^2-Q+p/\rho)=0$$

or, since $d(U+p/\rho)=dH$, where $H$ is the enthalpy per unit mass,

$$dQ=dH+VdV. \quad (44)$$

For a perfect gas, $dH=c_pdT$, whence the energy equation is finally

$$dQ=c_pdT+VdV. \quad (45)$$

### *Conservation of Mass*

The conservation of mass, in the form useful here, states that in a steady flow the mass entering a closed surface during any time interval $\Delta t$ is equal to the mass leaving during the interval $\Delta t$. Let the closed surface consist of the sections at arbitrary $\xi_1$ and $x_2$ (Fig. 1) and the portions of the duct between these sections, and let $\Delta t$ be the time required for the mass $m(\xi_1)-m(x_1)$ to enter the surface while mass $m(\xi_2)-m(x_2)$ flows out. The value of $x_1$ is arbitrary, whereas $\xi_2$ is fixed by the conditions on the time interval. Upon definition of $t(x)$ as the time required for a fluid particle to travel from origin $x=0$ to $x=x$, $\Delta t$ may be defined by

$$\Delta t \equiv t(\xi_1)-t(x_1)=t(\xi_2)-t(x_2). \quad (46)$$

The law of conservation of mass says that

$$\int_{x_1}^{\xi_1} \rho Adx = \int_{x_2}^{\xi_2} \rho Adx. \quad (47)$$

Upon change of variable, this equation becomes

$$\int_{t(x_1)}^{t(\xi_1)} \rho A\frac{dx}{dt}dt = \int_{t(x_2)}^{t(\xi_2)} \rho A\frac{dx}{dt}dt. \quad (48)$$

As before, by the theorem of the mean for integrals and condition (46), the integrand must be constant, whence

$$d(\rho VA)=0. \quad (49)$$

### *Equation of Motion*

The vector form of the second law of motion for continuous media states that the integral of the density of surface forces over a closed surface

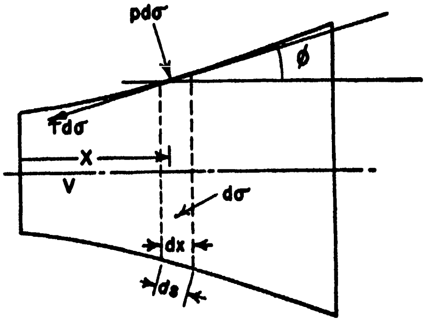


FIG. 2. Forces acting upon a fluid element described in the one-dimensional approximation.

is equal to the integral of the density times the particle derivative of the velocity over the volume enclosed by the surface.[25] For the mass of gas contained within the sections at $x_1$ and $x_2$ (Fig. 1) and the walls of the duct, the horizontal component of the equation of motion becomes

$$p(x_1)A(x_1)-p(x_2)A(x_2)+\int_{x_1}^{x_2} Rdx =\int_{x_1}^{x_2}\rho\frac{dV}{dt}Adx, \qquad (50)$$

where $R(x)dx$ is the horizontal component of the force exerted on the gas by the portion of the duct between $x$ and $x+dx$. For steady flow, the integral on the right-hand side may be transformed as follows:

$$\int_{x_1}^{x_2}\rho\frac{dV}{dt}Adx=\int_{x_1}^{x_2}\rho\frac{dV}{dx}\frac{dx}{dt}Adx=\rho VA\int_{x_1}^{x_2}\frac{dV}{dx}dx=\rho VA[V(x_2)-V(x_1)]. \qquad (51)$$

It follows from Eqs. (50) and (51) that the equation of motion may be written in differential form as

$$pdA+Adp+\rho VAdV=Rdx. \qquad (52)$$

Now $Rdx$ may be resolved into two constituent parts, the horizontal components of the tangential frictional drag and of the normal pressure reaction. If the half-angle of the duct is denoted by $\phi(x)$, the wall surface by $\sigma(x)$, and the tangential frictional drag per unit area by $\tau(x)$, from Fig. 2 it is clear that

$$Rdx=-(\tau d\sigma)\cos\phi+(pd\sigma)\sin\phi,$$

and that

$$d\sigma=dA/\sin\phi,$$

hence

$$Rdx=-(\tau d\sigma)\cos\phi+pdA. \qquad (53)$$

It is possible to use Eq. (53) directly without further analysis if the friction factor is related to $\tau$ by the equation $\tau=f\rho V^2/2$. In many engineering treatments, however, $f$ is defined in terms of the "energy loss due to friction." In order to make this concept of energy loss more precise and to make possible derivation of a rigorous connection between $\tau$ and energy loss, define $F(x)$ as the work done by unit mass of the fluid against friction in moving from the origin to position $x$. The work done in moving the entire mass of fluid between $x_1$ and $\xi_1$ to the region between $x_2$ and $\xi_2$ (see Fig. 3) will be computed in terms of the original variables $\tau$ and $\sigma$ and in terms of the new variables $F$ and $m$. If the two quantities are equated and suitably transformed, a relation will be obtained between $dF$ and $\tau d\sigma$.

Let $x_1$, $\xi_1$, and $x_2$ be chosen arbitrarily, and let $\xi_2$ be determined by the condition that the total mass $\Delta m$ between $x_1$ and $\xi_1$ equal the total mass between $x_2$ and $\xi_2$. Let $x_a$ be an arbitrary point between $x_1$ and $\xi_1$, and let $x_b$ be determined by the condition that $m(x_a)-m(x_1)=m(x_b)-m(x_2)$. In particular, if $x_a=x_1$, then $x_b=x_2$; if $x_a=\xi_1$, then $x_b=\xi_2$. Thus it is seen that, as $x_a$ runs from $x_1$ to $\xi_1$, $x_b$ runs from $x_2$ to $\xi_2$.

In order to determine the work done in terms of the variables $\tau$ and $\sigma$, the procedure is to move thin sections from their original positions, given in each case by $x_a$, to their final positions $x_b$, to find the work done by each of them, and then to add up the work done for all the sections.

First, divide the mass $\Delta m$ between $x_1$ and $\xi_1$ into smaller elements of mass $\Delta_1 m$, $\Delta_2 m$, $\cdots\Delta_n m$. (See Fig. 3.) Let $\Delta_i\sigma$ be the wall surface corresponding to the element of mass $\Delta_i m$, and let the duct wall between the initial and the final position of $\Delta_i m$ be divided into elements of length $\Delta_j s$. For a given $\Delta_i m$ it follows from Fig. 3

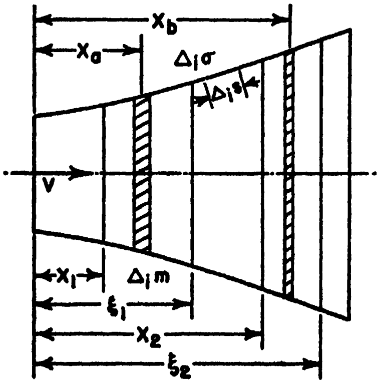


FIG. 3. Fluid elements considered in deriving the equation of momentum conservation.

---

[25] H. V. Craig, *Vector and Tensor Analysis* (McGraw-Hill Book Company, Inc., New York, 1943), pp. 355–356.

that

$$\Delta_i m = \rho_i A_i \Delta_i x,$$
$$\Delta_i\sigma = 2\pi r_i \Delta_i s,$$
$$\Delta_i x = \cos\phi_i \Delta_i s,$$

so that

$$\Delta_i\sigma = g_i \Delta_i m,$$

where $g_i \equiv 2\pi r_i/\rho_i A_i \cos\phi_i$. Each of the quantities $\rho_i$, $r_i$, $\phi_i$, and $g_i$ is to be evaluated at the proper point within $\Delta_i x$.

Finally, from Fig. 3 there exists on $\Delta_i m$ a force $\Delta F_i(s_j)$, to be overcome by the work against friction, equal to

$$\tau(s_j)\Delta_i\sigma(s_j) = \tau(s_j)g_i(s_j)\Delta_i m,$$

where it is convenient to consider $\tau$ and $\sigma$ as functions of $s$, inasmuch as frictional drag and the extension of an element of fixed mass depends on the position of the element in the tubes.

For any element of mass $\Delta_i m$ then, an approximation (the accuracy of which depends on the size of $\Delta_i m$) to the work done by $\Delta_i m$ in going from section $x_a$ to section $x_b$ is

$$\lim_{\Delta_j s\to 0} \sum_{s(x_a)}^{s(x_b)} \Delta F_i(s_j)\Delta_j s = \lim_{\substack{k\to\infty \\ \Delta_j s\to 0}} \sum_{j=1}^{k} \tau(s_j)g_i(s_j)\Delta_i m \Delta_j s$$
$$= \Delta_i m \int_{s[m(x_{a_i})]}^{s[m(x_{b_i})]} \tau(s)g_i(s)ds.$$

It follows that the corresponding approximation to the work done by the entire mass $\Delta m$ is

$$\sum_{i=1}^{n} \left\{ \int_{s[m(x_{a_i})]}^{s[m(x_{b_i})]} \tau(s)g_i(s)ds \right\} \Delta_i m$$

and, as the approximation becomes more and more accurate as the largest $\Delta_i m$ becomes smaller and smaller, the expression for the work done against friction when the entire mass between $x_1$ and $\xi_1$ is transported to the position between $x_2$ and $\xi_2$ becomes

$$\lim_{\substack{n\to\infty \\ \Delta_i m\to 0}} \sum_{i=1}^{n} \left\{ \int_{s[m(x_{a_i})]}^{s[m(x_{b_i})]} \tau(s)g_i(s)ds \right\} \Delta_i m$$
$$= \int_{m(x_1)}^{m(\xi_1)} \int_{s[m(x_a)]}^{s[m(x_b)]} \tau(s)g(s)dsdm. \quad (54)$$

It is clear that the work done against friction when the entire mass is moved as previously described is equal also to

$$\lim_{\substack{\Delta_j F\to 0 \\ \Delta_i m\to 0}} \sum_i \sum_j \Delta_j F \Delta_i m = \int_{m(x_1)}^{m(\xi_1)} \int_{F[m(x_a)]}^{F[m(x_b)]} dFdm$$
$$= \int_{m(x_1)}^{m(\xi_1)} \int_{F[m(x_a)]}^{F[m(x_b)]} (dF/ds)dsdm \quad (55)$$

whence it follows, since the intervals for both integrations are arbitrary, that

$$dF/ds = \tau g \quad (56)$$

and

$$dF = \frac{2\pi r}{\rho A \cos\phi} \frac{\tau d\sigma}{2\pi r}$$

whence

$$(\tau d\sigma)\cos\phi = \cos^2\phi \rho A dF. \quad (57)$$

Because, for the small angles usually under consideration, $\cos^2\phi$ is very nearly unity (for a half-angle of 6°, $\cos^2\phi = 0.989$), the retention of this factor except for particularly precise work would not seem justifiable. Hence Eq. (53) may be written

$$Rdx = -\rho AdF + pdA. \quad (58)$$

The differential equation of motion is finally

$$-dp = \rho VdV + \rho dF. \quad (59)$$

The connection between $dF$ and the differential loss in stagnation pressure $(-dp_t)$ can, with the help of Eqs. (14) and (15), be expressed in the form

$$\frac{-dp_t}{p_t} = \frac{dF}{RT} + \frac{\gamma N d\theta}{[2+(\gamma-1)N]}. \quad (60)$$

Thus except for the limiting case of incompressible flow, $(-dp_t)$ and $dF$ cannot be used interchangeably in defining the friction factor even for the adiabatic case where $d\theta = 0$.

## APPENDIX B

### The *Z* Language

In the place of $N = V^2/\gamma RT$ the equations may be formulated in terms of $Z$, defined through

$$Z = (V^2/2c_p)/(T + V^2/2c_p). \quad (61)$$

The numerator is the so-called dynamic temperature, the denominator is the total temperature. $Z$ and $N$ are related by

$$Z=[(\gamma-1)N]/[2+(\gamma-1)N], \quad (62)$$

$$N=2Z/(\gamma-1)(1-Z). \quad (63)$$

This replacement represents a one-to-one transformation of $N$ into $Z$ in the range 0 to infinity for $N$, 0 to 1 for $Z$.

In order to illustrate the form that some of the earlier equations take under this transformation, Eqs. (14)–(16) are written in terms of $Z$,

$$\frac{dZ}{Z}=\frac{2}{1-[(\gamma+1)/(\gamma-1)]Z}\left\{\frac{1-Z}{2}\left[1+\left(\frac{\gamma+1}{\gamma-1}\right)Z\right]d\theta+(1-Z)d\mu+(1-Z)d\alpha\right\}, \quad (64)$$

$$\frac{dp}{p}=\frac{2}{1-[(\gamma+1)/(\gamma-1)]Z}\left(-\frac{\gamma}{\gamma-1}Zd\theta-\frac{1+Z}{2}d\mu-\frac{\gamma}{\gamma-1}Zd\alpha\right), \quad (65)$$

$$\frac{dT}{T}=\frac{2}{1-[(\gamma+1)/(\gamma-1)]Z}\left[\left(\frac{1-Z}{2}-\frac{\gamma Z}{\gamma-1}\right)d\theta-Zd\mu-Zd\alpha\right]. \quad (66)$$

---

# Crystalline Aggregation of Cobalt Powder[a]

J. T. McCartney[b] and R. B. Anderson[c]

(Received April 28, 1947)

In electron microscopic studies of Fischer-Tropsch catalysts, an interesting phenomenon was observed in cobalt metal powder reduced from cobaltous oxide. The oxide particles sintered into larger smooth droplets of cobalt that were aggregated into thin hexagonal-shaped platelets. X-ray diffraction analysis showed the presence of the hexagonal crystal phase of cobalt. A possible explanation is that the aggregates are formed by forces similar to those operating in normal crystal formation but of reduced magnitude insufficient to destroy the identity of the component particles.

In the course of examination, with the electron microscope, of various components of cobalt catalysts used in the Fischer-Tropsch synthesis of liquid and solid hydrocarbons from mixtures of carbon monoxide and hydrogen, an interesting phenomenon was observed in cobalt metal powder reduced from the oxide. The cobalt was prepared by precipitation of cobaltous oxide from cobaltous nitrate with aqueous ammonia, reduction of the oxide in hydrogen at 250°C, and stabilization of the metallic cobalt against rapid oxidation by exposure to carbon dioxide at liquid nitrogen temperature. Figures 1 and 2 are electron micrographs showing the relative sizes and shapes of the particles of the oxide and the reduced metal. Surface area measurements by a nitrogen adsorption method[1] gave values of 67 square meters per gram for the oxide and 2 square meters per gram for the reduced cobalt. Thus, during reduction, there was an evident sintering of the irregular particles of the oxide into larger smooth droplets of cobalt metal.

Such droplets have been observed in several other metals prepared and reduced in a similar way. In cobalt, however, these particles are more or less closely aggregated into hexagonal platelets such as shown in Figs. 3 and 4. Varying degrees of dispersion in different specimens are illustrated in Figs. 2–4. The hexagons vary in diameter from 7 to 25 microns, and their thickness is apparently of the order of the diameter of the component particles. The hexagonal shape can be discerned also in the optical microscope. X-ray diffraction analysis shows that the crystal phase present is hexagonal cobalt. The hexagons are quite stable,

---

[a] Contribution from the Central Experiment Station, Pittsburgh, Pennsylvania. Published by permission of the Director, Bureau of Mines, U. S. Department of the Interior.

[b] J. T. McCartney, Physicist, Coal Constitution and Miscellaneous Analysis Section, Bureau of Mines, Pittsburgh, Pennsylvania.

[c] R. B. Anderson, Physical Chemist, Office of Synthetic Liquid Fuels, Research and Development Division, Bureau of Mines, Pittsburgh, Pennsylvania.

[1] S. Brunauer, P. H. Emmett, and E. Teller, J. Am. Chem. Soc. 60, 309 (1938).

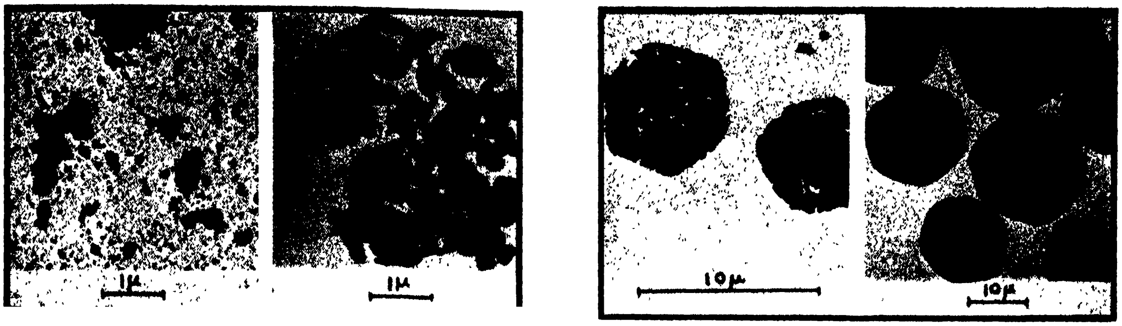


FIG. 1. FIG. 2. FIG. 3. FIG. 4.

FIGS. 1–4. Electron micrographs of cobaltous oxide (Fig. 1) and of cobalt metal reduced from it, in various stages of dispersion (Figs. 2–4).

their presence being observed in reduced cobalt samples which were at least partly oxidized and in reduced cobalt partially converted to cobalt carbide.

A possibly related phenomenon is that described by Picard and Duffendack[2] in evaporated metals. Their electron micrographs of evaporated zinc and cadmium show hexagonal crystals of the order of 0.1 micron in diameter. These are explained as resulting from migration and combination of the condensed atoms of the metals. In the Bureau of Mines laboratory, cobalt was evaporated onto collodion substrates and observed in the electron microscope. It apparently formed nearly continuous films with no observable migration into large crystals. Such was the case even at elevated temperatures, the cobalt condensing onto evaporated silica films heated probably above 250°C. Since cobalt has a much higher melting point than zinc and cadmium, higher temperatures may be required for migration.

It is suggested that the hexagonal aggregates in cobalt are formed by forces similar to those operating in normal crystal formation. Since, however, the component particles do not lose their identity and the aggregates are easily fractured, they can hardly be classed as true crystals. Possibly the forces acting to combine atoms into crystals are so reduced when the combining units are particles of relatively large size that only a superficial crystal formation occurs. It is not evident why this phenomenon has been observed only in cobalt.

---

[2] R. G. Picard and O. S. Duffendack, J. App. Phys. **14**, 291 (1943).

# Surface Studies of Glass.

## Part I. Contact Angles

L. A. Spitze* and D. O. Richards
*Physics Laboratory, Owens-Corning Fiberglas Corporation, Newark, Ohio*

(Received May 5, 1947)

The water-repellent nature of glass surfaces treated with different silicones is examined by means of contact-angle measurements. Results are given showing the effect of various concentrations of silicones, types of glass, temperature of heat treatment, length of time of heat treatment, and surface roughness on the water contact angles. Silicones containing the methyl and phenyl group show good heat stability at temperatures of 500°C for short periods of time.

### INTRODUCTION

In applying glass to many industrial uses, more information is desired concerning its surface properties. This information is increasingly important as the surface area-weight ratio becomes progressively larger. In Fig. 1, the relationship between surface area and diameter of fibers is shown in English and metric units. From this graph it can be seen that fibers of glass are now being made having a surface area more than 50,000 times that of an equal unit weight of bulk glass. These extended surfaces present an opportunity for varied physical and chemical reactions to occur, and emphasize the need for an increasing amount of work on surface studies of glass. Such surface phenomena may possibly alter or mask many of the basic properties of the glass itself with the resultant properties being largely influenced by the type of surface presented. To learn more about some of these properties, different physical characteristics of the surface can be studied. A property of prime consideration is the degree of wetting of glass surfaces by liquids. Investigations of the degree of wetting of glass surfaces by liquids can be done readily by measurements of the contact angles these liquids form on the glass surface. Harkins and others[1] have shown that the contact angle for water, benzene, carbon tetrachloride, chloroform, and ether in contact with clean glass is zero. A solid is completely wetted by a liquid if the contact angle is zero, and incompletely wetted if a measurable contact angle is found. A zero contact angle indicates that the liquid attracts the solid as much as it attracts itself, a 90° contact angle shows that the attraction of the liquid for the solid is half that for itself, and a contact angle of 180° would mean there was absolutely no attraction of the liquid for the solid. Because contact angles of 180° are unrealizable, complete non-wetting is impossible; but solids with liquid contact angles greater than 90° are generally referred to as being non-wetted. In such industrial applications of glass fibers as flotation mats, water-repellent fabrics, and glass-plastic laminates with decreased water sorption, a hydrophobic surface is desired. Such materials as the long-chain fatty acids or their salts, when applied to a glass surface, will make it hydrophobic.. Although these materials will give a hydrophobic surface, they possess poor durability under various conditions, and particularly poor stability at elevated temperatures. Recently a new group of materials has been introduced containing members that also have the ability to give a hydrophobic surface to glass when treated by them. These materials, commonly referred to as the silicones,[2] have the distinct advantage of good heat stability. Materials such as the silcones, when applied to glass surfaces, permit the use of much higher temperatures in the applied product than would be available for glass surfaces treated with the usual organic materials. Thus, more of the inherent heat-resistant properties of the glass will be available for use in products containing glass

---

* Present address: Department of Chemistry, Southwestern College, Winfield Kansas.

[1] Harkins and Brown, J. Am. Chem. Soc. 41, 499 (1919); Richards and Carver, J. Am. Chem. Soc. 46, 1196 (1924).

[2] C. A. Scarlott, Westinghouse Engineer 5, 130 (1945); H. Hausman, J. Chem. Ed. 23, 16 (1946); H. A. Knight, Materials and Methods 23, 1069 (1945); S. L. Bass, Elec. Eng. 66, 381 (1947).

treated with these materials. In addition, these silicones, being largely inorganic materials containing silicon, could be expected to be compatible with a clean glass surface and would firmly adhere to such surfaces.

The degree of non-wetting that the silicones impart to a glass surface depends on the individual member, and on the method by which the silicone is applied to the glass surface.

## EXPERIMENTAL

The hydrophobic nature of a glass surface treated with various silicones was studied by determining the water contact angles. Measurements of contact angles may be made by various procedures. Methods most commonly used include direct measurement of the angle by use of photography or projection, tilting plate,[3] sessile drop,[4] or pressure of displacement in capillary tubes. The method selected for this study was the sessile drop. Determinations are made by placing a number of small drops of water on the treated glass surface and then measuring the dimensions of the drops with the aid of a microscope. The size of the contact angle, $\theta$, is then calculated from the relationship[5] $\tan\theta/2 = 2h/d$ where $h$ is the height of the drop, and $d$ is the width of the drop at the place of contact with the solid. The sessile drop method can be used only if the portion of the drop measured is a segment of a sphere. It has been shown that if the size of the drops is kept smaller than 1 mm in diameter, the contact angle attains a constant maximum value that is the advancing contact angle.[4] This method was selected for this study because of the ease of adapting the available equipment, and measurements of a number of drops on a treated surface will give an average value for the drops placed at various positions on the treated surface.

The apparatus used to determine the contact angles of water on the treated glass surfaces consisted of a Bausch and Lomb microscope placed in a horizontal position and containing a 10× objective. The drop was viewed and measured by means of a Bausch and Lomb 12.5× Filar eyepiece. Most of the glass samples used were cut to about 1″×½″ by ¼″ thick. The smooth, uncut, fire-polished upper surface was used in the studies. The treated glass sample was placed on a stage and brought into the line of sight by a mechanical microscope stage. A Spencer microscope illuminator was used to give a sharp outline of the drops placed on the glass surface. Drops of distilled water of fairly constant size, less than 1 mm in diameter, were placed on the treated glass surfaces by means of capillary tubes. For some of the treated surfaces having high contact angles, it became necessary to place the drops on the surface by spraying, as the adhesion tension of the water to the glass capillary was much greater than that for the treated glass surface. No fewer than five separate drops in the desired size range were measured and contact angles calculated. The glass samples were first cleaned with chromic acid cleaning

TABLE I. Contact angles of water on trichloroorganosilane treated *E*-glass.

| 15 Min. @ °C | 30° | 75° | 125° | 200° | 300° | 400° | 500° |
|---|---|---|---|---|---|---|---|
| Trichloromethylsilane | 97 | 99 | 104 | 103 | 101 | 96 | 94 |
| Trichlorophenylsilane | 81 | 93 | 100 | 101 | 97 | 91 | 84 |
| Trichloroethylsilane | 92 | 100 | 103 | 97 | 90 | 60 | 0 |
| Butyltrichlorosilane | 96 | 98 | 101 | 96 | 90 | 60 | — |
| Trichlorooctylsilane | 98 | 99 | 101 | 96 | 86 | 0 | — |
| Trichlorododecylsilane | 111 | 108 | 107 | 103 | 84 | 35 | 0 |
| Trichlorooctadecylsilane | 106 | 106 | 103 | 70 | 66 | 10 | 0 |

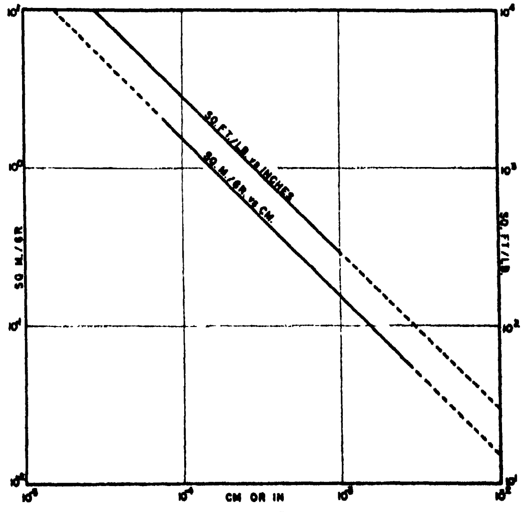


FIG. 1. Surface area *vs.* fiber diameter; solid lines show present working range in the fibrous-glass industry.

---

[3] N. K. Adam and G. J. Jessop, J. Chem. Soc. (London) **127**, 1863 (1925); E. L. Green, J. Phys. Chem. **33**, 921 (1929); R. N. Wenzel, Ind. Eng. Chem. **28**, 988 (1936).

[4] G. L. Mack, J. Phys. Chem. **40**, 159 (1935).

[5] F. E. Bartell, *Laboratory Manual of Colloid and Surface Chemistry* (Edwards Brothers, Inc., Ann Arbor, Michigan, 1938), p. 85.

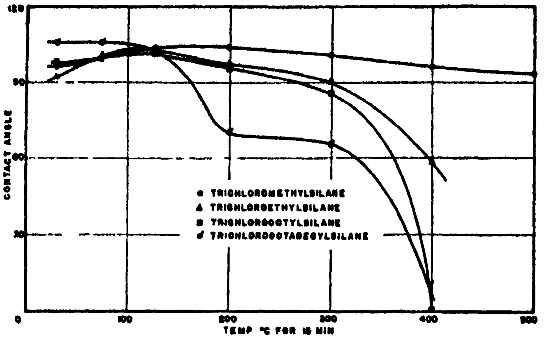


FIG. 2a. Contact angles of water on $E$-glass treated with 1 percent solution of trichloroorganosilanes *vs.* temperature.

solution and then by flaming. The cleaned glass was then dipped in a 1 percent solution of the silicone in c.p. benzene, removed and allowed to drain a short while, and then placed in an oven at the desired temperature. After the sample was in the oven the desired time, it was removed and allowed to cool. The treated samples were then placed on the microscope stage and the dimensions of the drops placed on them were measured by use of the Filar eyepiece in the microscope.

## RESULTS

Values of the contact angles of water were obtained for glass surfaces treated with organosilane chlorides and with organosiloxane polymers. The magnitude of the contact angle was found as a function of the variables of temperature, length of time of heat-treatment, type of glass, and concentration of the surface treatment.

### Trichloroorganosilanes

Various members of this particular group of silicones were used to treat the glass surfaces. These materials were diluted with c.p. benzene until a solution was obtained containing 1 percent solids. Contact angles were obtained for these materials, first as a function of the temperature, the other variables being held constant. The samples of glass used here were a borosilicate glass type containing almost no alkali ($Na_2O$ or $K_2O$) and designated as $E$-glass by Owens-Corning Fiberglas Corporation. After being treated with the silicone solutions, the samples were placed in the oven at the desired temperature and left for fifteen minutes. The results of these measurements are shown in Table I. The data of Table I have also been presented graphically in Figs. 2a and 2b. These values represent determinations made at different times on different samples of the same type of glass. At temperatures below 300°C the results were most consistent. At the higher temperatures, the results varied somewhat. This can be expected from the shape of the curves as shown in Figs. 2a and 2b. Because the curves are steep in this region, any small variation in heating temperature or length of time of heating would greatly affect the results.

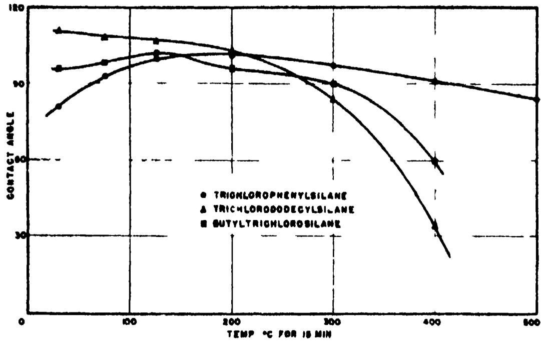


FIG. 2b. Contact angles of water on $E$-glass treated with 1 percent solution of trichloroorganosilanes *vs.* temperature.

TABLE II. Contact angles of water on $E$-glass treated with organosiloxane polymers.

| 15 Min. @ °C | 30° | 75° | 125° | 200° | 300° | 400° | 500° | 550° |
|---|---|---|---|---|---|---|---|---|
| Dimethylsiloxane polymer | 31 | 45 | 67 | 92 | 101 | 99 | 96 | 31 |
| Diethylsiloxane polymer | 40 | 46 | 82 | 102 | 97 | 91 | 44 | 25 |
| Methylphenylsiloxane 0.5 Ph/Si | 31 | 43 | 49 | 90 | 101 | 97 | 97 | 80 |
| Methylphenylsiloxane 0.9 Ph/Si | 25 | 44 | 61 | 91 | 91 | 90 | 91 | 91 |

The effect of the type of glass being treated on the contact angle is shown in Fig. 3. Three different types of glasses were used, treated with trichlorooctadecylsilane (1 percent solids), and heated at the specified temperatures for fifteen minutes. The three glass types used were $E$-glass, Pyrex chemical glass, and window glass. Chemical analysis shows that these glasses contain different amounts of alkali ($Na_2O$ and $K_2O$). The $E$-glass has less than 1 percent, the Pyrex chemical glass about 4½ percent, and the window glass about 12½ percent of these alkali materials.

## Organosiloxane Polymers

In Table II are shown the results of contact-angle measurements of water on *E*-glass treated with 1 percent solutions in benzene of four different organosiloxane polymers. These materials were heated at the desired temperature for fifteen minutes. The data have been plotted in Fig. 4 for visual use of these results.

It was considered advisable to determine the effect of different periods of heating on the contact angles of glass surfaces treated with these polymers. In Table III is shown the effect of various heating periods up to six hours on *E*-glass. Three different temperatures were used for this study: room temperature, 300°C, and 450°C. An upper temperature of 500°C was tried at the beginning, but for periods greater than one-half hour the results varied widely. Consequently, this "burning off" region was avoided and a lower temperature of 450°C selected. The results obtained are also shown graphically in Figs. 5–7.

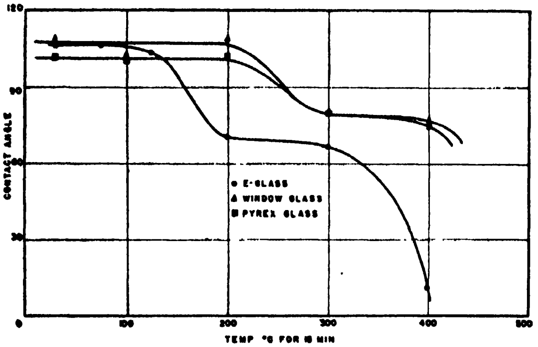


FIG. 3. Contact angles of water on various glasses treated with 1 percent solution of trichlorooctadecylsilane *vs.* temperature.

In addition, the effect of various concentrations of the silicones on *E*-glass was studied by contact-angle measurements. A curve showing the effect of dimethylsiloxane polymer on *E*-glass heated at 200°C and at 500°C for 15 minutes is shown in Fig. 8. A trichloroorganosilane was also used to study its effect at different concentrations. The results obtained for this material are discussed later.

### DISCUSSION

TABLE III. Contact angles of water on *E*-glass treated with organosiloxane polymers.

| Material | Temp. °C | Heating time in minutes 15 | 30 | 60 | 180 | 360 |
|---|---|---|---|---|---|---|
| Dimethylsiloxane polymer | 30 | 31 | 37 | 38 | 40 | 46 |
| | 300 | 101 | 104 | 104 | 100 | — |
| | 450 | 101 | 93 | 93 | 81 | 81 |
| Diethylsiloxane polymer | 30 | 40 | 53 | 54 | 55 | 55 |
| | 300 | 97 | 101 | 93 | 87 | — |
| | 450 | 89 | 79 | 62 | 62 | 20 |
| Methylphenylsiloxane 0.5 Ph/Si | 30 | 38 | 34 | 35 | 36 | 34 |
| | 300 | 101 | 96 | 99 | 96 | — |
| | 450 | 93 | 99 | 90 | 82 | 56 |
| Methylphenylsiloxane 0.9 Ph/Si | 30 | 25 | 45 | 39 | 39 | 36 |
| | 300 | 91 | 90 | 96 | 93 | — |
| | 450 | 96 | 99 | 93 | 79 | 86 |

Before making contact-angle measurements by the sessile drop method on silicone-treated glass, the method was tested by comparing the results obtained on some common materials with the values of the contact angles for these materials as reported in the literature. Measurements made of the contact angle of water on paraffin gave an average value of 105°, which compares favorably with the values reported elsewhere.[6] A photo-micrograph of a drop of water on paraffin is shown in Fig. 9. From this photograph it can be noted that the drop is a segment of a sphere, and the sessile drop method of measurements will give the contact angle for drops this size.

Two additional photo-micrographs of water on *E*-glass treated with dimethylsiloxane polymer are shown in Figs. 10 and 11. Figure 10 shows the appearance of water on *E*-glass treated with the polymer and heated to 125°C for 15 minutes. After the treated glass has been heated to 300°C for 15 minutes, the appearance is shown in Fig.

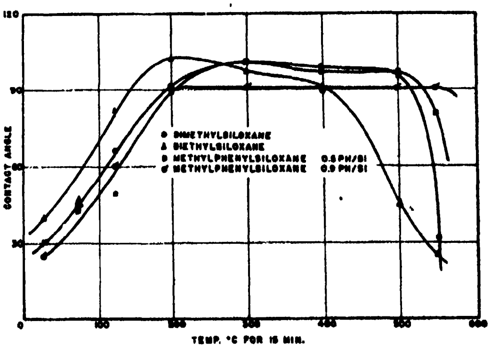


FIG. 4. Contact angles of water on *E*-glass treated with 1 percent solution of organosiloxane poymers *vs.* temperature.

---

[6] Int. Crit. Tab. 4, p. 434 (1928); N. K. Adams, *The Physics and Chemistry of Surfaces* (Oxford University Press, London, 1941), p. 186.

11. This increase of heat raises the contact angle 34°.

Observation of the data in Table I and the curves of Figs. 2a and 2b bring out the superior heat stability of the trichloroorganosilanes containing the methyl and the phenyl groups. These data show that the heat stability as determined by water repellency of these trichloroorganosilanes is dependent upon the particular organic group attached to the silicon. From these results it appears that the heat stability of the organic groups tested may be arranged in this descending order: methyl, phenyl, ethyl, butyl, octyl, dodecyl, and octadecyl. The maximum water repellency is shown by the materials containing the dodecyl and the octadecyl groups. Although the data in Table I do not show a definite order, it is most likely that the arrangement of the organic groups in the order of decreasing water repellency would be the reverse of the order for heat stability as given above. In particular, the methyl group seems to deviate from this arrangement The maximum contact angles shown by these materials are all within a 10° range; hence any experimental errors may account for the fact that the data are not in the orderly arrangement suggested above. The data for the organosiloxane polymers again show the heat stability of the methyl and the phenyl group. For these materials, the polymer containing the highest ratio of phenyl to silicon shows the best heat stability. Although there is very little difference in the values for the maximum contact angles obtained, the polymer containing the ethyl group appears to have a slightly higher maximum contact angle.

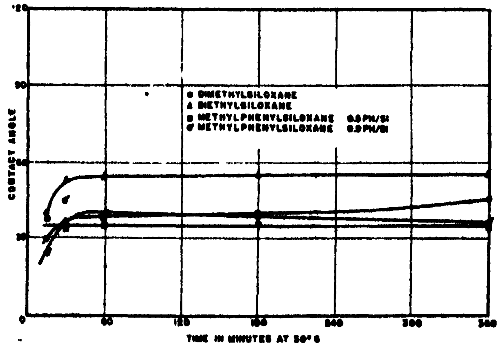


FIG. 5. Contact angles of water on *E*-glass treated with 1 percent solution of organosiloxane polymers *vs*. heating time.

The curves of Fig. 3 show a noticeable difference in the contact angles of various types of glasses treated with trichlorooctadecylsilane. These curves show that the highest contact angles are obtained for the glass containing the most alkali. Thus it is believed that the higher contact angles for window glass and for Pyrex glasses are due to the higher alkali content. It appears that the silicone materials will attach themselves more firmly to the alkali glasses, possibly because of a higher surface activity of these glasses. Results obtained when a dimethylsiloxane polymer was used to treat these various glasses did not show such a difference as exhibited by the trichlorooctadecylsilane-treated glasses. The two treated glasses containing the most alkali did, however, show appreciably higher contact angles at treatment temperatures below 200°C and values a little higher at temperatures above 500°C. It is possible that the polymer is able to orientate itself somewhat faster on these glasses containing alkali and that it is also bound a little stronger to these glasses.

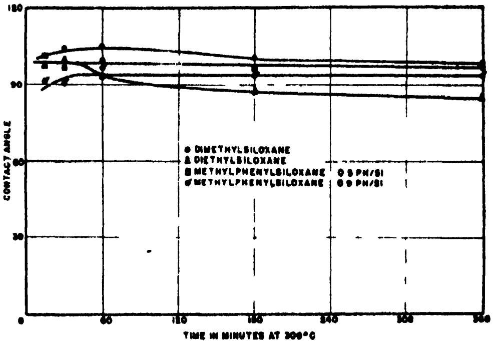


FIG. 6. Contact angles of water on *E*-glass treated with 1 percent solution of organosiloxane polymers *vs*. heating time.

The mechanical condition of the surface of the treated glasses influences the wettability to a large extent. The effect of the roughness of a surface on the contact angles has been recognized[7] as due to a greater net energy change in the wetting process caused by a rough surface. A drop tending to spread on a rough surface will increase the free liquid surface, but the wetted area under the drop will increase to a larger extent because of the rough surface. Increase of

---

[7] R. N. Wenzel, Ind. Eng. Chem. 28, 988 (1936).

the liquid surface involves a consumption of energy, while the increase of wetted area under the drop involves a release of energy. The difference is a net energy change which determines the speed of wetting on the solid. For a rough solid tending to be wet by a liquid, a greater net energy decrease is present than if the surface were smooth and this causes the rough surface to be wetted more thoroughly. For a liquid-repellent rough surface, the same reasoning can be applied. For a change in the shape of a drop on a rough liquid-repellent surface, more actual surface change is involved at the liquid-solid interface. As the dry interface has the lower specific energy, the net energy decrease is greater for the rough surface causing the drop to assume a more spherical form. A roughened surface magnifies the wetting properties of a solid, making it either more liquid-repellent or more easily wetted.

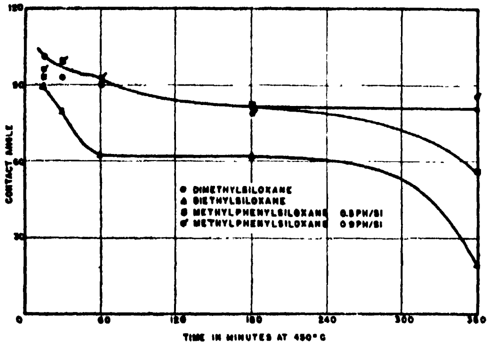


FIG. 7. Contact angles of water on *E*-glass treated with 1 percent solution of organosiloxane polymers *vs.* heating time.

The effect of a roughened glass surface on the contact angles was investigated briefly. A sample of *E*-glass was roughened by grinding with an abrasive and another sample was roughened with an acid etch. These surfaces were cleaned and then treated with a 1 percent solution of dimethylsiloxane polymer and heated for 15 minutes at 300°C. The contact angles obtained on these samples gave values of from 120° to 125°, depending somewhat on the manner by which the surface was roughened. A photo-micrograph of the appearance of a drop of water on a roughened *E*-glass surface treated with dimethylsiloxane polymer at 300°C for 15 minutes is shown in Fig. 12. The contact angle for this sample was found to be 124°, or an increase of 23° over the value of a treated smooth surface. The industrial applications of the surface roughness of materials, as an aid to water-repellency, are being studied and evaluated at the present time.

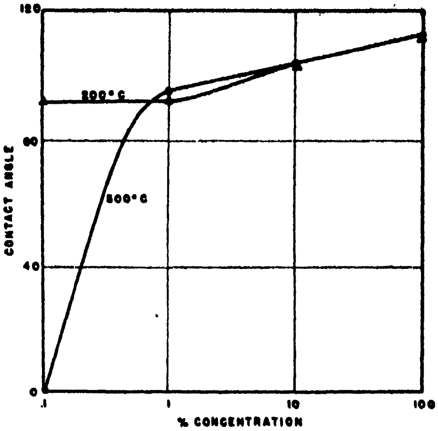


FIG. 8. Contact angles of water on *E*-glass treated with dimethylsiloxane polymer at 200°C and 500°C *vs.* concentration.

The effect of various lengths of time of heating these treated glass surfaces is small at room temperature. The data in Table III and Figs. 5–7 show, also, that at 300°C there is a small effect, if any, because of variation in length of time of heating these treated samples. Two of the samples appear to have a slightly larger contact angle as the time is increased to 30 minutes, but the effect is small. At 450°C, a more noticeable effect begins to appear. At this higher temperature, some of the material is beginning to burn off. This is particularly so for the diethylsiloxane polymer, and to a lesser extent for the methylphenylsiloxane polymer (0.5 Ph/Si). The data for the diethylsiloxane polymer also show a drop in contact angle with time up to 60 minutes, then remain steady until it begins to burn off. This drop may be due to the fact that equilibrium for this material is not reached at the higher temperatures until after 60 minutes. This effect is also somewhat noticeable for the dimethylsiloxane polymer.

The effect of different concentrations of the silicones on glass is significant, particularly at the lower concentrations. The data in Fig. 8 show the effects of varying the concentration of

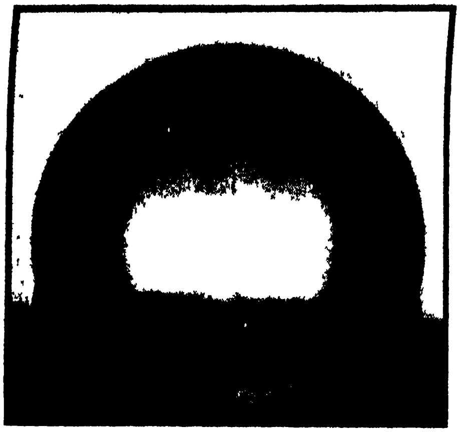


FIG. 9. Appearance of a drop of water on a paraffin surface; contact angle 105°.

dimethylsiloxane on *E*-glass and then heating at 200° and at 500°C for 15 minutes. The data show that an increase in concentration raises the contact angle at a particular temperature. This is most pronounced at the higher temperature and the lower concentration. At 500°C the 0.1 percent solution burns off, but the 1 percent solution gives a high contact angle. The 0.1 percent solution, although heated at 200°C, is not destroyed and gives a high angle. Although experiments were not tried at lower concentrations, it is believed that lower concentrations of dimethylsiloxane polymer would yield lower contact angles for the 200°C curve. Thus if these experiments had been conducted at lower concentrations, the resultant contact angle *vs.* temperature curves would very probably have been lower and burned off at considerably lower temperatures. The increase of the contact angles resulting from increase of concentration may be due in part, at the higher temperatures, to an increased surface roughness. This roughness is probably caused by deposition of some $SiO_2$ on the glass by the high heat, and any undestroyed silicone will give the necessary water-repellency necessary to obtain the higher contact angles. The concentration effect in the case of the trichloroorganosilanes is not as pronounced as that of the organosiloxane polymers. However, at the higher temperatures and lower concentrations, a noticeable drop in the values for the contact angles was apparent. These materials would also be expected to give values of contact angles somewhat lower as the concentration is decreased, but not to such an extent as that shown by the organosiloxane polymers.

Water-repellent glass in the form of fine fibers has proved to be very useful. For example, these treated glass fibers have proved their superiority over other materials as buoyant materials for Navy life preservers.[8] This superiority is due to their fire- and mildew-proofness, superior buoyancy, ability to withstand repeated wetting and drying with little loss in efficiency, and the fact that the treated glass fibers can be produced in

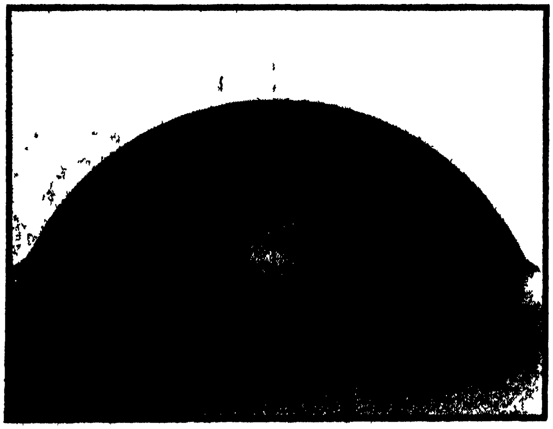


FIG. 10. Appearance of a drop of water on an *E*-glass surface treated with a 1 percent solution of dimethylsiloxane polymer and heated for 15 minutes at 125°C; contact angle 67°.

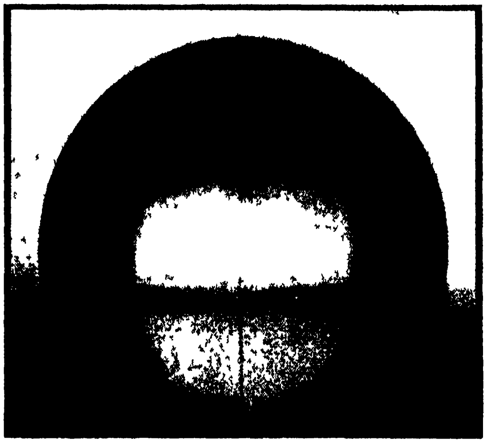


FIG. 11. Appearance of a drop of water on an *E*-glass surface treated with a 1 percent solution of dimethylsiloxane polymer and heated for 15 minutes at 300°C; contact angle 101°.

---

[8] A. L. Holden, U.S.N. Inst. Proc. 72, 1327 (1946).

uniform quality. Such treated glass fibers have been adopted by the Navy as standard material.[8] In addition, these silicone-treated glass fibers have shown their value in reinforced plastics.[9] Their water-repellent properties have been very useful in increasing wet strength of these laminates. The wet strength of cloth laminates, improved by the application of a silicone solution to the glass fibers, has been brought up to a point nearly equal to the dry strength.[10]

This paper shows the results of contact-angle measurements of only a few of the many possible types of silicone materials. More information obtained on the other types of silicone materials available, when applied to glass and other materials, will be of utmost importance. Studies other than contact-angle measurements will also prove their worth in evaluation of surface properties of surfaces treated by the silicones.

## CONCLUSIONS

The values of the contact angles of water on glass surfaces treated with various trichloroorganosilanes and organosiloxane polymers show the good water-repellency properties these materials bestow on a glass surface. In addition, the superior heat stability of these materials is demonstrated. Of the various organic groups in the silicone molecule, the methyl and phenyl impart the best heat-resistant properties, and the dodecyl and octadecyl give the materials their best water-repellency. The value of increased surface roughness in giving better water-repellency also is shown. Finally, the effect of variations in the concentrations of these silicones when applied to the surfaces is given. It is recommended that concentrations no lower than 1 percent solids be used when maximum heat stability and water-repellency is desired.

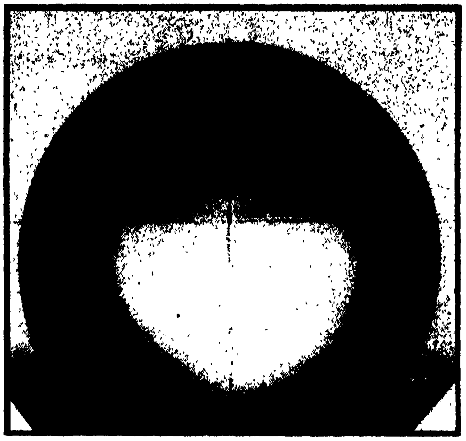


FIG. 12. Appearance of a drop of water on a roughened surface of *E*-glass treated with 1 percent solution of dimethylsiloxane polymer and heated for 15 minutes at 300°C; contact angle 124°.

## ACKNOWLEDGMENT

The silicone materials used in this study were furnished by Dr. M. J. Hunter of Dow-Corning Corporation, Midland, Michigan. The authors also wish to acknowledge the aid given during the course of this study by R. B. Taylor, W. C. Marlow, L. P. Biefeld, Robert Steinman, and G. F. Forman, all of the Research Division of Owens-Corning Fiberglas Corporation.

---

[9] P. L. Alger, Gen. Elec. Rev. 50, 12 (1947).

[10] G. F. Nordenholt, Prod. Eng. 18, 160 (1947).

# An Apparatus for the Observation of the Kerr Effect with Microsecond Electrical Pulses

WILBUR KAYE AND RICHARD DEVANEY
*Tennessee Eastman Corporation, Research Laboratories, Kingsport, Tennessee*

(Received May 9, 1947)

An apparatus for the observation of the Kerr electro-optic effect in liquids and polymers using microsecond electrical pulses, is presented and evaluated. This method offers a convenient way of studying internal motions and molecular configurations of polymeric material with a minimum of heating usually accompanying these observations. Electrical pulses of from 1- to 1000-microsecond width, 60–1000 p.p.s., at amplitudes up to 5000 volts are generated by conventional methods. These pulses are used to orient low molecular weight and polymeric materials placed between the plates of a condenser. The circular polarization induced in the polarized light passing between the plates actuates a photo-multiplier tube in order that the light pulses may be studied with an oscillograph.

Curves of Kerr constant *versus* temperature at temperatures to −80°C are compared with the literature. Oscillograms showing the response of several liquids and polymers are shown. Oscillograms showing an interesting difference between orientation and relaxation time are presented, as are oscillograms showing clearly both orientation and distortion polarization simultaneously.

BRIEFLY the Kerr electro-optic effect is concerned with the elliptical polarization induced in a material subjected to an electrical field.[1] It is the aim of this paper to present the instrumental details and the evaluations necessary to utilize this Kerr effect for the study of internal motions and molecular configurations of organic liquids and polymers.

The instrument described in this report utilizes pulsed electrical fields for the study of the Kerr effect. This method offers the advantage of minimizing heating effects in the material under investigation, and permits one to study the nature of the orienting effect in a more direct manner than has hitherto been accomplished by the use of sinusoidal electrical fields.[2] A circuit similar to that used here was described by Pauthenier[3] using pulsed condenser discharges. However, Pauthenier had no direct method of studying the shape of the response wave as is now practical by use of the oscillograph.

Figure 1 shows a photograph of the exterior of the apparatus, which in conjunction with Fig. 2, a block diagram, will be used to explain the methods by which the Kerr effect may be produced and observed.

Electrical pulses of the order of a few microseconds width are generated every hundredth of a second by the pulse generator. The high voltage amplifier then boosts the voltage of the pulses to several thousand volts amplitude without serious distortion. This electrical pulse on the plates of the Kerr cell causes the plane-polarized light passing through the sample to be circularly polarized; hence, it passes through the analyzing Nicol which is crossed with respect to the polarizing Nicol and oriented 45° with respect to the field in the Kerr cell. The light pulse subsequently produced by the electrical field actuates the photo-tube, wherein it is converted into an electrical signal that can be amplified and studied with the oscillograph.

Ordinary low molecular weight compounds are capable of orienting in an electrical field within $10^{-9}$ second;[4] hence, the wave shape of the optical pulse is very nearly the same as that of the electrical pulse, provided certain restrictions of wave form and amplitude are made. On the other hand, there are materials composed of large molecules in a viscous medium which do not follow the electrical field at such a rapid rate. The optical pulse from such material should be expected to show considerable distortion in relation to the electrical field if one assumes a molecular orientation effect to be responsible for the electro-optic effect.

---

[1] J. Kerr, Phil. Mag. 50, 337, 446 (1875).
[2] J. W. Beams, Rev. Mod. Phys. 4, 133 (1932).
[3] M. J. Pauthenier, J. de phys. et rad. 2, 183, 384 (1921).
[4] J. W. Beams and Lawrence, J. Frank. Inst. 206, 109 (1928).

The amount of light passing the analyzing Nicol is related to the electrical field intensity, $F$, and the optical path length, $l$, between the Kerr cell plates by the expression[5]

$$I = a^2 \sin^2(\pi K l F^2) = a^2 \sin^2 \delta/2, \tag{1}$$

where $K$ is the Kerr constant and $\delta$ the angular retardation. The general relationship of $K$ with other molecular factors is[6]

$$K = \frac{(n^2-1)(n^2+2)}{4n^2} \frac{(\epsilon+2)^2}{(3)} \frac{\theta_1+\theta_2}{\alpha}, \tag{2}$$

where $n$ is the refractive index, $\epsilon$, the dielectric constant, $\alpha$, the polarizability, and $\theta_1$ and $\theta_2$ are defined as

$$\begin{aligned} \theta_1 = (1/45kT)[&(a_1-a_2)(b_1-b_2) \\ &+(a_2-a_3)(b_2-b_3)+(a_3-a_1)(b_3-b_1)], \\ \theta_2 = (1/45k^2T^2)[&(\mu_1^2-\mu_2^2)(b_1-b_2) \\ &+(\mu_2^2-\mu_3^2)(b_2-b_3)+(\mu_3^2-\mu_1^2)(b_3-b_1)]. \end{aligned}$$

Here, $k$ is Boltzmann's constant; $T$, the absolute temperature; $a_1$, $a_2$, and $a_3$ are proportional to the shape of the induced polarization ellipsoid of the molecule in three mutually perpendicular directions; $b_1$, $b_2$, and $b_3$ are similar coefficients of the optical ellipsoid of the molecule; while $\mu_1$, $\mu_2$, and $\mu_3$ are the components of the permanent dipole moment. The importance of the temperature dependance arising from thermal motion will be discussed later.

From what has been said above, it should appear possible to find a material which would show a definite lag at easily obtainable temperatures in response to the electrical field, and that lag would be markedly dependent on temperature. Such a material would have to be composed of large polar molecules in a viscous medium. These requirements are met by a number of synthetic plasticizers and polymers. The literature records the observation of such lags in rosin,[7] glass,[8] benzopurpurin,[9] and vanadium pentoxide solutions.[10]

## GENERATION OF PULSE

In this section use will be made of Fig. 3, a schematic diagram of the pulse amplifier, to give a brief description of its operation. The pulse generator is a Measurements Corporation model 79-B unit. The output circuit of this generator has been modified somewhat to give control of amplitude and output impedance as well as modified to obtain wide pulses. Normally, the generator is used with full amplitude (approximately 150 v) and 1000-ohm output impedance. Pulses from ½ to 1000 microseconds and repetition rates from 60 to 100,000 p.p.s. are obtainable.

The output coupling condenser has been removed for certain applications to be mentioned later. However, the coupling condenser is in the circuit located on the amplifier chassis to permit control of pulse amplitude through the grid bias. Such a method of amplitude control was used to permit the design of twin amplifiers using a common signal source and plate supply. The amplifier tubes are Eimac 4-125A tubes. Two pairs of load resistors are available, a 10,500-ohm load for narrow pulses and a 75,000-ohm load for wide pulses.

Taps in the load resistors permit a portion of the output signal to be deflected directly to the plates of the cathode-ray tube for observation of

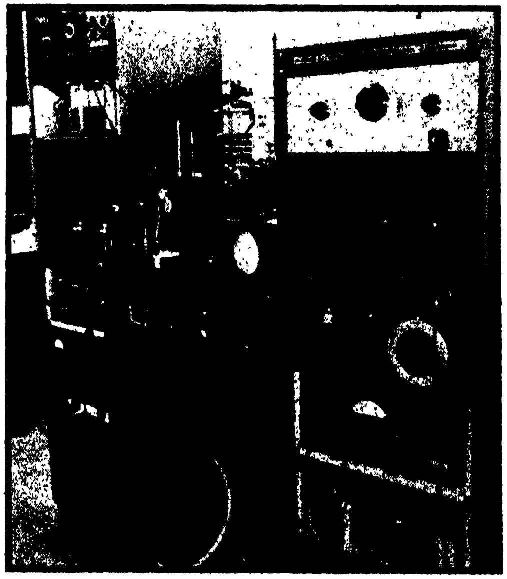


FIG. 1. Exterior of apparatus. Pulse generator in left rack, optical bench, amplifier, and oscillograph on bench and dielectric bridge in right rack.

---

[5] E. F. Kingsbury, Rev. Sci. Inst. 1, 27 (1930).

[6] P. Debye, Handbuch Radiologie 6, 754 (1925).

[7] J. Kerr, see reference 1; D. W. Kitchin and H. Muller, Phys. Rev. 32, 979 (1928).

[8] J. Kerr, see reference 1.

[9] J. Errera, J. Overbeek, and H. Sack, J. de Chimie Phys. 32, 681 (1935).

[10] J. Errera, J. Overbeek, and H. Sack, see reference 9.

the electrical wave form. This circuit attenuates the output signal but not the a.c. ripple; hence, it greatly enhances this ripple.

Another tap in one of the load resistors permits the use on the Kerr cell of an attenuated pulse, improving the wave form when both channels are operated simultaneously, one with considerably greater amplitude than the other.

The peak pulse voltage is measured by means of a Sensitive Research Instrument Corporation model ESH multi-range electrostatic voltmeter. The pulses are isolated by means of a diode as shown. Care has to be taken in this circuit to place a high resistance to ground to prevent the accumulation of a d.c. charge formation through leakage of the coupling capacitor. The two amplifier channels are carefully shielded for accurate voltage determinations. Switching is accomplished by means of Eimac vacuum switches in the high voltage circuits and antenna relays in the low voltage circuits.

Figure 4 is an oscillogram of the electrical pulse of 10-microseconds width. The wave form has been improved by driving the circuit to saturation in the sense that the pulse voltage approximates the plate supply.

### OPTICAL SYSTEM

The light source mounted on a Cenco optical bench is a Spencer 6-volt lamp operated from a battery to avoid any periodic fluctuations of light intensity. Cenco 10-mm Nicol prisms act as polarizing and analyzing media. The Kerr cells are mounted ridgidly in line. Figure 5 shows the construction of one of these cells and its temperature bath which has operated successfully to keep the temperature constant to within one degree from −80 to +100°C, using a dry ice ether mixture in the side chambers and an immersion heater with De Kotinsky thermo-regulator to maintain the temperature in an alcohol or oil medium. The light enters and leaves through double windows maintained clear at low temperatures by a stream of air. Considerable difficulty was encountered before suitable cell plates were found which could be kept parallel at separations less than a millimeter and still be small and easily cleaned after use with polymeric materials. A solution to this problem was found by coating ordinary glass microscope slides with a layer of gold deposited from duPont Liquid Bright Gold and fired at 1100°F for a short time. Both sides of the slides are coated, as are three of the edges, and electrical contact is made by spring contacts to the outside surface of the slides. The slides are separated by mica spacers averaging 0.016 in. Thicker glass spacers have been used for materials of high Kerr constant.

A Gaertner model L-135 Babinet-Soleil compensator is placed between the Nicols to permit measurement of the retardation, zero out small retardations in the glass windows, and increase the sensitivity of the system as the occasion may require. The compensator requires 4.42 turns per wave of light of 465-mu wave-length.

The analyzing Nicol is crossed with respect to the polarizing Nicol and at an angle of 45° with respect to the plane of the electrical field. A Baird interference filter passing light in the neighborhood of 456 mu was placed directly over the iris diaphragm acting as a window to the photo-tube chamber.

The type 931-A multiplier photo-tube was chosen because of its sensitivity and frequency response.[11] The voltage-divider system for the

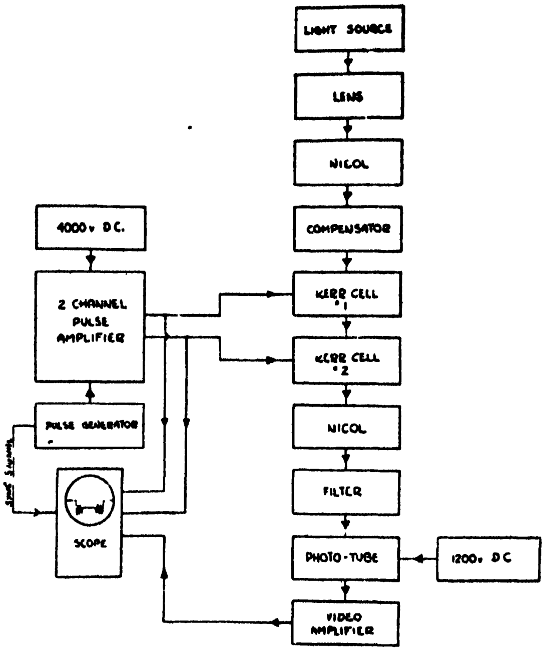


FIG. 2. Block diagram of Kerr effect apparatus.

---

[11] R. C. Winans and J. R. Pierce, Rev. Sci. Inst. 12, 269 (1941); R. D. Rawcliffe, Rev. Sci. Inst. 13, 413 (1942).

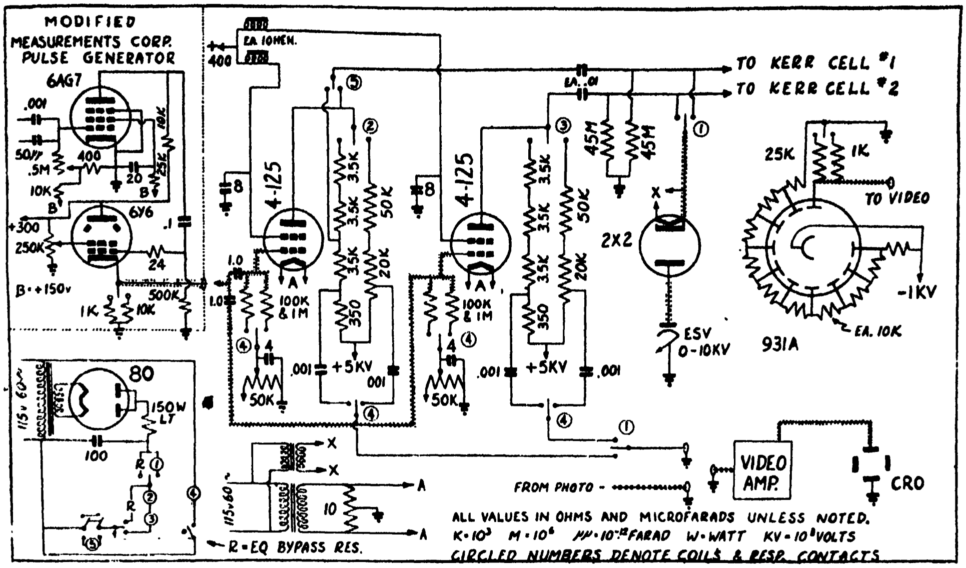

MODIFIED MEASUREMENTS CORP. PULSE GENERATOR
6AG7
6Y6
80
4-125
4-125
2X2
931A
TO KERR CELL #1
TO KERR CELL #2
TO VIDEO
-1KV
ESV 0-10KV
+5KV
+5KV
VIDEO AMP.
CRO
FROM PHOTO
R-EQ BYPASS RES.
115v 60~
ALL VALUES IN OHMS AND MICROFARADS UNLESS NOTED.
K=10³ M=10⁶ μμ=10⁻¹²FARAD W=WATT KV=10³VOLTS
CIRCLED NUMBERS DENOTE COILS & RESP. CONTACTS

FIG. 3. Wiring diagram of Kerr effect apparatus.

multiplier dynodes is mounted in a shielded box directly below the photo-tube. The load resistor of the photo-tube is variable to permit control of the output amplitude and fidelity of the high frequency response. This resistor likewise is mounted close to the photo-tube and connected to the video amplifier by a short shielded lead.

The video amplifier is a United Cinephone model 8240 with a gain of 1000 and a frequency essentially flat from 15 cycles to 4 megacycles. The output of this video amplifier feeds directly to the plates of the cathode-ray tube in a DuMont 208B oscilloscope.

## INSTRUMENTAL RESPONSE

An elaborate discussion of the optics of this type of system will not be given here since it may be found in the literature.[12] However, attention should be given to formula (1), applicable to the above conditions where it may be seen that the light intensity, hence deflection of the oscilloscope tracing, will be primarily a function of the field strength. With a material of sufficiently large Kerr constants it is possible to observe a sinusoidal fluctuation of light intensity on the leading and trailing edges of the pulse. Figure 6 shows an oscillogram of several superimposed optical pulses with nitrobenzene at low voltage. A comparison of this figure with that of the electrical pulse will furnish a measure of the performance of the instrument. Figure 7 is an oscillogram of the response of nitrobenzene to a large voltage pulse. For this photograph the electrical circuit was set to produce sloping edges and high amplitude. With reference to Eq. (1) it is seen that seven waves of retardation are produced. In order to obtain such a pattern, it is necessary to use monochromatic light and an amplifier of high fidelity in addition to accurately parallel cell plates properly aligned.

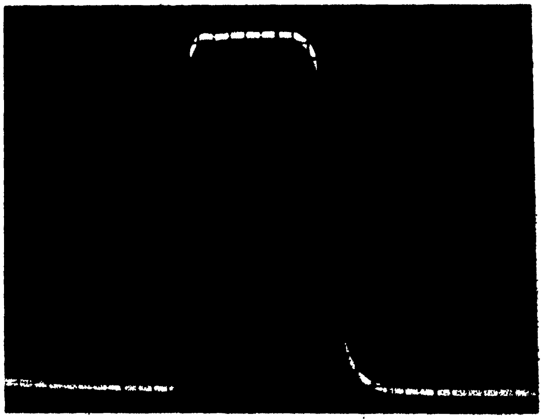


FIG. 4. Oscillogram of electrical pulse; 10-μsec. width, 650 p.p.s. 4500-volts amplitude.

---

[12] E. F. Kingsbury, see reference 5.

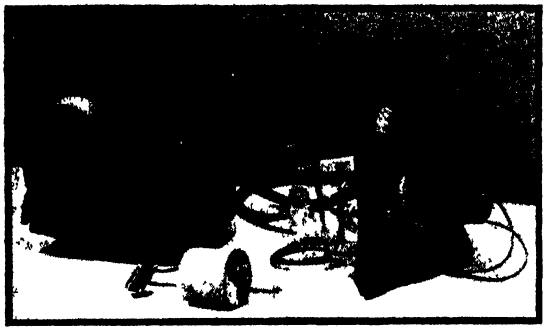


FIG. 5. Kerr cell and temperature bath.

## METHODS OF MEASUREMENT

The simplest method of measuring the Kerr constant would be to measure the amplitude of the signal and compare this amplitude with a material of known Kerr constant. This method, however, is inaccurate though useful for the estimation on materials of very low Kerr constant.

A method of greater accuracy uses the Babinet-Soleil compensator to produce a retardation equal and opposite to that produced in the Kerr cell. The procedure is as follows.

The amplitude of the pulsed retardations produced by the Kerr cell is set to an arbitrary value by varying the gain of the video amplifier. The compensator is introduced and set to a value that will just *invert* the signal an equal and opposite amount. The difference in compensator readings from its null point (previously determined) to the point giving the above retardation is divided by the number of turns per wave to give the waves of angular retardation equal to that produced by the pulse. As noted in Eq. (1), the angular retardation is proportional to the square of the field strength, the constant of proportionality being in turn proportional to the Kerr constant. Hence, a plot of the retardation *versus* the field strength squared will result in a straight line through the origin when everything is properly aligned. If such a graph be obtained for two materials, one of which is $CS_2$ of known Kerr constant and the other the material in question, the desired Kerr constant will be

$$K_x = K_{CS_2}(\delta_x/F^2/\delta_{CS_2}/F^2).$$

This procedure, of course, presumes the use of the same Kerr cell for both materials.

It should be emphasized that in the determination of the retardation by this method the compensator retardation was varied until the signal on the oscilloscope screen inverted an equal and opposite amount. It must be remembered that the photo-tube amplifier circuit does not interpret d.c. signals; hence, the light at the inverted point is striking the photo-tube at all times except during the pulse. This very fact handicaps this method of measurement for the photo-tube fatigues under conditions of constant light as well as increases the noise level. Both of these effects are clearly visible and introduce appreciable errors.

If a second Kerr cell is used in addition to the compensator, the matter of photo-tube fatigue may be minimized. The plates of the Kerr cells must be so oriented as to give retardations of opposite sign (90° rotation if the Kerr constants are of like sign, 180°, if of unlike sign). Using synchronized pulses on the two Kerr cell plates, the voltages on the cells may be adjusted to nullify each other. Figure 8 is an oscillogram of such a null point with $CS_2$ and $CHCl_3$. Only the leading and trailing edges are apparent above the noise level of the photo-tube. The Kerr constant for an unknown relative to $CS_2$ will then be

$$K_{CS_2}{}' = C \cdot K_{CHCl_3}{}'\left(\frac{V_{CHCl_3}}{V_{CS_2}}\right)^2,$$

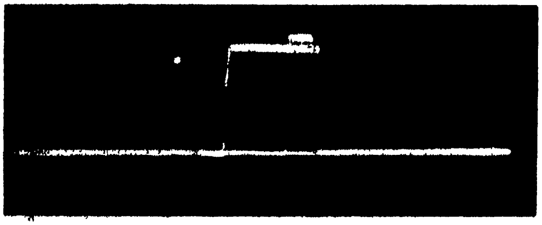


FIG. 6. Oscillogram of optical pulse with nitrobenzene with weak yield at three pulse widths 10, 40, and 150 μsec. 60 p.p.s. 1900 volts, glass separator.

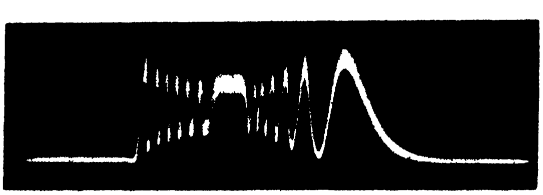


FIG. 7. Oscillogram of nitrobenzene response to high field strength, 28-μsec. width, 400 p.p.s. amplitude 3180 volts with mica spacer, light of 4500A.

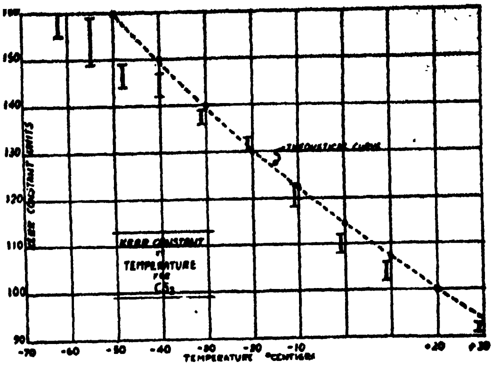


FIG. 9. Kerr constant *vs.* temperature for carbon disulfide.

where $C$ is a constant accounting for differences in cell spacing and length. Figure 9 shows a graph of the Kerr constant *versus* temperature for $CS_2$ determined in this manner, with $CHCl_3$ being used as a standard for null purposes. Changes in temperature ($t'$) of the $CHCl_3$ must be accounted for accurately. The agreement with the literature is good.[13] A small amount of constant retardation was introduced by the compensator. This increases the sensitivity of the system without changing the null voltage ratio. Differentiation of Eq. (1) will show the reason for this.[14] No filter was used in determining the curve in Fig. 9, but the data will apply to the average wave-length of approximately 500 mu.

## APPLICATIONS

An application which was mentioned previously as of importance is the study of relaxation in polymers. The molecular orientation hypothesis for the explanation of the Kerr effect as well as usual dielectric absorption effects has been fairly well established;[15] hence, we may picture the Kerr effect as arising from the orientation of molecules or portions of molecules upon the application of an external electrical field. If the molecule is large, and in a viscous medium, the time of orientation may be large. The apparatus described here should be well suited for the observation of this relaxation phenomenon being, in a sense, more direct than the conventional dielectric methods.

Figure 10 is an oscillogram of this effect in Santolite MS, a highly viscous plasticizer, taken at 56°C. At this temperature the molecules or dipoles completely orient in approximately 40 microseconds. However, the relaxation seems to proceed more rapidly than the orientation—an observation which was not expected. At a slightly higher temperature (65°C) the molecules are capable of orienting in 5–10 microseconds as can be seen in Fig. 11. Upon further increase in temperature the orientation becomes more rapid, but the amplitude of the effect (Kerr constant) decreases as one should expect from Eq. (2). If the temperature is lowered and the pulse width increased, it is possible to observe both the orientation and distortion polarizations simultaneously. Figure 12 shows this effect at room temperature with an 800-microsecond pulse. The increased noise level is a result of high amplifier gain. The optical pulse rises very rapidly to a value considerably less than in Fig. 11, then slowly increases until the end of the pulse when it immediately returns to an intermediate level and tapers off slowly to zero. The distortion polarization, of course, accounts for the sharp rise and decline, while the slow change may be attributed to the orientation effect. By such methods it should be possible to isolate and measure the theta-terms of Eq. (2). The dis-

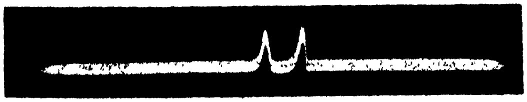


FIG. 8. Typical null point using two-cell technique chloroform and carbon disulfide 10-μsec. width, 300 p.p.s., room temperature.

FIG. 10. Relaxation in Santolite MS 40-μsec. pulse, 300 p.p.s. 3900 volts, temperature 56°C.

---

[13] A. A. Zuchlka and L. R. Ingersol, J. Opt. Soc. Am. 27, 314 (1937).

[14] E. F. Kingsbury, see reference 5.

[15] J. W. Beams, see reference 2.

FIG. 11. Relaxation in Santolite MS 20-μsec. pulse, 300 p.p.s. 4000 volts, temperature 63.5°C.

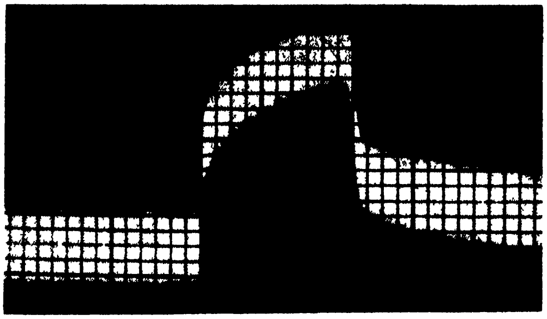


FIG. 12. Oscillogram of Santolite MS at 15°C, 800-μsec. width, 60 p.p.s., 4150 volts; note immediate distortion polarization followed by gradual orientation.

persion region, measured by conventional dielectric methods, lies in the same range of temperature for the conventional dielectric method as for the electro-optic, giving strong indication to the fact that the same unit (molecule or dipole) is orienting in both cases.

This relaxation effect in polymers can be observed dielectrically with pulsed fields by placing a test condenser of the dielectric directly across the load resistor of the pulse generator cathode-follower circuit and observing the relaxation with a cathode-ray tube. The effect is not as pronounced, because of other instrumental effects, but has been observed to be temperature and time dependent in the same manner as in the Kerr effect.

Further studies are in progress correlating this Kerr effect with conventional dielectric properties and viscosity for a series of primary alcohols.

# On the Localized Afterglows Observed with Long Sparks in Various Gases

J. D. CRAGGS AND W. HOPWOOD
*Research Department, Metropolitan-Vickers Electrical Company, Ltd., Manchester, England*

AND

J. M. MEEK
*Department of Electrical Engineering, Liverpool University, Liverpool, England*

(Received June 2, 1947)

Simultaneous photographs, taken in two perpendicular directions, have been obtained of long sparks in various gases at 1-atmosphere pressure. The experiments were undertaken in order to decide whether the localized afterglows, noticed as long streaks in rotating mirror spark photographs, were caused by some discharge mechanism or by purely geometrical effects resulting from the tortuous nature of the spark channels. The results show that the second explanation is not feasible. Some qualitative interpretations of the effects are given.

## I. INTRODUCTION

THE study of spark discharges in gases has usually been confined to the examination of such discharges in air and to the earlier parts of the process, e.g., the formation of streamers.[1] Recently the later phases of spark formation (spark channels) have been re-examined and some of their properties have been described by Raiskiy (1940),[2] Flowers (1943),[3] and by Craggs and Meek (1946),[4] following the earlier work of several authors such as Lawrence and Dunnington (1930),[5] Ollendorff (1933),[6] and others. Such spark channels may be considered, in brief, to consist of a highly conducting path (i.e., of low voltage gradient) in which the electron temperature may be 10,000–20,000°, and approximating in some aspects to a nascent arc of artificially restricted diameter. Rotating mirror camera records of lightning, taken mainly by Schonland and his collaborators to whom much of the best work is due, and of laboratory sparks (Allibone and Meek (1938), (a) and (b),[7,8] and others) have shown, however, that the last stages of a spark discharge, perhaps even when the current flow has sensibly ceased (Craggs and Meek)[4] are represented by localized afterglows. The significance of the latter has often been overlooked or minimized, and the purpose of the present paper is to present an account of further preliminary experiments made on this effect.

## II. DISCUSSION OF PREVIOUS EXPERIMENTS

Many previous papers, e.g., those of Allibone and Meek[7,8] tend to ignore the afterglow streaks, shown clearly in a typical record such as Fig. 1 (b)(ii) of Allibone and Meek's second paper[8] and concentrate on other aspects of the discharge. Walter (1935),[9] in a discussion on Hoffert's early photograph (1889)[10] of a multiple lightning stroke, noticed the streaks and stated that "they are by no means due to an afterglow of the lightning channel produced by thermal or phosphorescent causes, but they are always associated with a real after-discharge in the channel, i.e., with an electric current following the main discharge along the same track." Walter considered that this conclusion is justified by the fact that the streaks often increase in brightness at an appreciable time after the light from the main channel has greatly decayed. An excellent example of this effect has been published by Allibone and Meek[8] (Fig. 1 (a)(iii)); in such cases there is almost certainly an appreciable oscillatory flow of current, but such currents should be constant at all points along the dis-

---

[1] L. B. Loeb and J. M. Meek, *The Mechanism of the Electric Spark* (Stanford University Press, Stanford University, California, 1941).

[2] S. M. Raiskiy, J. Tech. Phys. U.S.S.R. **10**, 529 (1940)

[3] J. W. Flowers, Phys. Rev. **64**, 225 (1943).

[4] J. D. Craggs and J. M. Meek, Proc. Roy. Soc. **A186**, 241 (1946).

[5] E. O. Lawrence and F. G. Dunnington, Phys. Rev. **35**, 396 (1930).

[6] F. Ollendorff, Archiv f. Elektrotechnik. **27**, 169 (1933).

[7] T. E. Allibone and J. M. Meek, Proc. Roy. Soc. **A166**, 97 (1938).

[8] T. E. Allibone and J. M. Meek, Proc. Roy. Soc. **A169**, 246 (1938).

[9] B. Walter, Phil. Mag. **20**, 1144 (1935).

[10] H. H. Hoffert, Phil. Mag. **28**, 106 (1889).

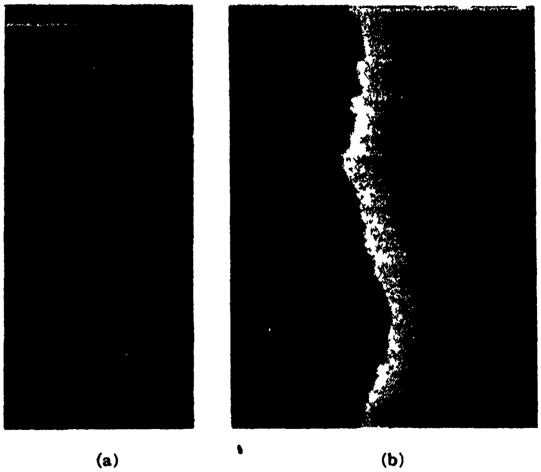


FIG. 1. Rotating film photographs of the point/plane sparks in air (1-atmos. pressure); (a) had 100-cm gap and a high circuit resistance; (b) had 130-cm gap and a low circuit resistance and is an enlarged record.

charge, a supposition which was apparently overlooked by Walter and which unless other effects are present would demand uniform afterglows along the channel at any particular instant. It is therefore very unlikely that "follow currents" can account for the localized streaks of afterglow, although accentuation of the latter by the former probably occurs.

McEachron and McMorris[11] appear to accept Walter's view in a modified form and suggest that very small (<100 amp.) follow currents may be responsible, but again ignore the peculiar localized nature of the afterglows. It is appropriate to state here that the measurement of small follow currents is difficult, and their existence in many laboratory experiments cannot be denied. Such measurements are now projected in this laboratory, although the arguments to be advanced below tend, it is suggested, to minimize their fundamental importance. The lightning photographs of Malan and Collens,[12] especially their Figs. 13–17 include some excellent records, and those authors agree with the conclusion of McEachron and McMorris.

Flowers[13] subscribes to an alternative theory of the afterglow streaks and states that "portions of the path (of the discharge) which are projected parallel to the direction of motion of the film or the lens provide a greatly increased exposure time and proportionate reduction in resolution. Path components which lie along the photographic axis increase the light intensity. Both effects produce excessive exposures and are associated with the streaks that are observed in nearly all Boys camera photographs." It is clear that, whilst Flowers does not apparently in this communication attribute the streaks directly or wholly to the effects described, which he states "are associated" with them, the effects of bends could contribute to or produce highly localized streaks. This explanation had occurred independently to us some years ago, but was immediately rejected, since the length of the afterglows seem to be of quite a different order of length from the average effective length of a bend in the channel.[14] It appeared necessary to disprove this theory, if possible, before embarking on further experiments since, if true, the bend

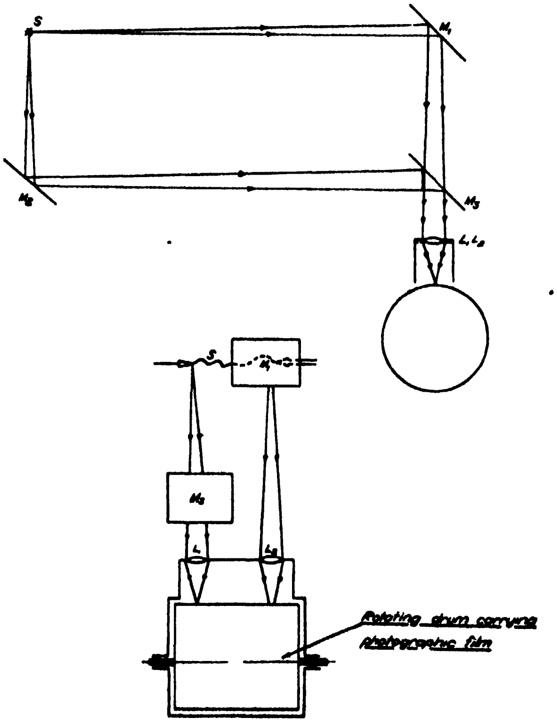


FIG. 2. Optical system used in right-angle spark photography; $M_1$, $M_2$, and $M_3$ are mirrors; $S$ is the spark discharge; $L_1$ and $L_2$ are lenses.

---

[11] K. B. McEachron and W. A. McMorris, Gen. Elec. Rev. 39, 487 (1936).

[12] D. J. Malan and H. Collens, Proc. Roy. Soc. A162, 175 (1937).

[13] J. W. Flowers, Gen. Elec. Rev. 47, 9 (1944).

[14] J. D. Craggs, Quart. J. Roy. Met. Soc. 70, 167 (1944) (discussion on paper by T. E. Allibone).

theory would entirely remove the physical importance of the effect. It is important to note that Flowers' suggestion refers to bends in the main channel and not in the much dimmer and more transitory leader strokes. (It does in fact appear from many records (e.g., those of Fig. 1) that the streaks arise in the leader strokes or streamers and persist through the later stages of the spark channel's formation and extinction.)

In the discussion following a paper by Allibone (1938),[15] O'Dea made the suggestion that the streaks may be due to the persistence of "ionized or ozonized pockets." Allibone, in reply, referred to the theoretical work of Neugebauer[16] on ball lightning. The present authors favor O'Dea's proposition, with the important correction that the presence of free electrons and ions, demanded by Neugebauer's theory, may not be necessary since the afterglows may be explained on other grounds. (Craggs and Meek,[4] with references to earlier work from 1943 by the same authors.)

As mentioned above, it appeared necessary first to investigate the possibility of a geometrical explanation of the afterglow streaks, i.e., the effect of sustained illumination of the photographic film caused by the increased exposures at channel bends, either in or perpendicular to the plane of the film. This paper describes such experiments.

## III. THE EXPERIMENTAL METHOD

The most obvious method for assessing the importance of the geometrical bend theory, outlined above, is to photograph the sparks with a rotating film or mirror camera in two different, preferably perpendicular, directions at once. If the results should, for example, show that the luminous afterglow streaks are noticed at points where no channel bending is perceptible, then it would appear reasonable to discount the bend theory and to suppose that some effect in the ionized or excited gas in the spark channel is operative. Further investigations to elucidate the nature of the effect would then be warranted.

Strigel[17] is apparently the only author to have photographed sparks previously in this way, but his object was not to investigate the afterglow streaks, which are indeed not mentioned in his paper, and it is not possible to use Strigel's reproductions of spark photographs in deciding the point at issue in the present paper.

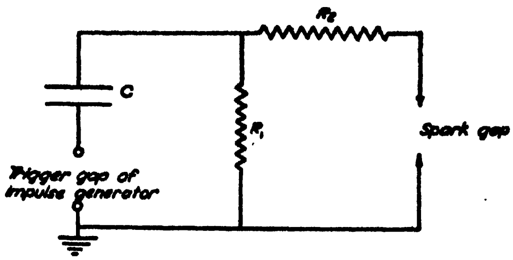


FIG. 3. Impulse generator; $C=6$ series stages of 0.22 uf per stage. Total discharge capacity$=0.037$ uf; $R_2=210$ ohms; $R_1=35{,}000$ ohms; charging voltage$=70$ kv.

The optical system is shown in Fig. 2, in which the rotating drum (3000 r.p.m.) carries the photographic film. The drum diameter was approximately 12″. The records (see Fig. 5) were therefore composed of pairs of photographs arranged one above the other on the film, each pair corresponding to one spark.

The spark was produced with a 2-million volt impulse generator (0.11 mfd/stage) rearranged to give effectively the circuit of Fig. 3, in which a peak current of 2000 amp. was used. The point/point gap of 20 cm needed some 280 kv for breakdown in air. Other experiments were made at 1000 and 500 amp., with suitable alterations to the circuit of Fig. 3. A current oscillogram for 2000-amp. peak current is given in Fig. 4.

For gases other than air, a glass vacuum chamber (30-cm diameter, 60 cm long) was used, pumped down before gas admission to a pressure of about 1 mm Hg. The gases from cylinders were then introduced to a pressure of 80 cm Hg, in order to prevent contamination by air leakage. It was found to be necessary to attach a black cloth inside the glass cylinder to prevent reflections. The thickness of the glass chamber was about 1 cm so that ultraviolet radiations from the sparks were prevented from reaching the film.

About 30 films, each with 10–20 pairs of records, were taken.

## IV. RESULTS OF THE EXPERIMENTS

Figure 5 shows some results (full-size reproductions of original negatives), obtained with 2000-amp. discharges in oxygen (3 sparks), hydro-

---

[15] T. E. Allibone, J. Inst. Elec. Eng. 82, 487 (1938).
[16] T. Neugebauer, Zeits. f. Physik 106, 474 (1937).
[17] R. Strigel, Wiss. Veröff. aus den Siemens Werken. 15, Part 3, 13 (1936).

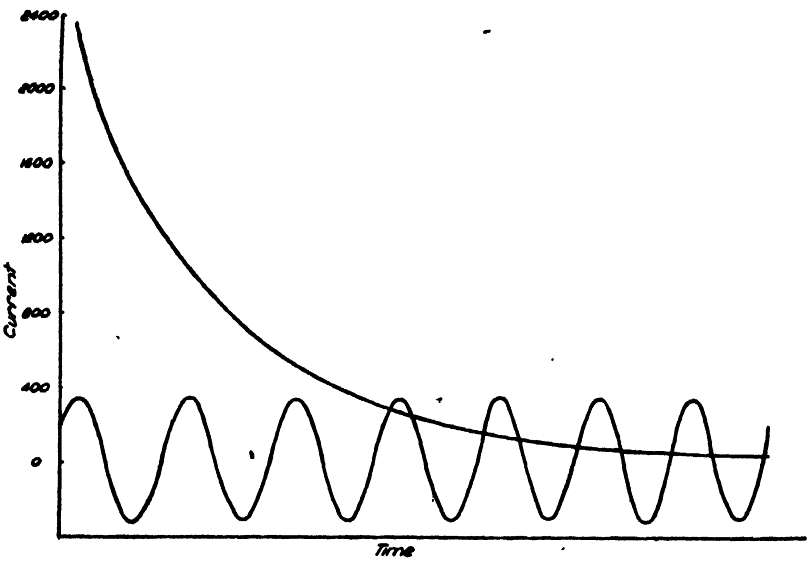


FIG. 4. 2000-amp. current oscillogram for spark discharge in argon. The timing oscillation has a period of 5 microsecs.

gen (2 sparks), and nitrogen (2 sparks) which are typical of the many other photographs taken (1 cm on the time axis ≡ 240 microseconds). The value of the method for assessing the importance of channel bends is at once apparent. Oxygen and nitrogen show pronounced afterglow streaks which are highly localized, particularly in nitrogen. The formation of branched channels was found with oxygen and nitrogen but not with hydrogen, and the latter gas showed very short afterglows, in agreement with previous experiments. The spark channels in hydrogen have a pronounced tendency to form bends which are particularly sudden and well defined. The very short afterglows in hydrogen are fairly localized, as shown by the tufted appearance of the channels on their afterglow side.

It is clear, especially from the oxygen records of Fig. 5 (a) that the streaks are formed in both records at the same points in the channel, not necessarily where bends or inflections are noticed, and are generally of the same length. These findings are in general disagreement with the geometrical theory. Where channel division occurs (Fig. 5 (a)) the different views of the spark may be different largely because of the geometrical (not time) separation of the channel branches with subsequent overlapping and confusion of the afterglows. The exceptionally tortuous bends in hydrogens do not give enhanced streaks, as would be expected from the geometrical theory which does not invoke any difference in channel structure from gas to gas.

When the argon results, typical specimens of which are given in Fig. 6, are examined it seems even clearer that the afterglow streaks are associated with localized discharge conditions in the channel. Figure 6 (b) particularly demonstrates the evidence; argon spark channels show only a tendency to form bends which are, even then, very gradual and yet the streaks are particularly long and well defined, and situated clearly at the same points and for the same times, in both records for each spark. Figure 6 (a) shows an exceptionally straight spark. Figure 6 (b) demonstrates another effect noticed in argon, i.e., the formation, at approximately the same point in the channel even for different sparks, of a single well-defined streak. This phenomenon is probably, it is thought, connected with the earlier parts of the discharge process and if substantiated this conclusion would constitute irrefutable evidence against the supposed importance of channel bends. It is, however, considered that the data given in Figs. 5 and 6 and described above constitute sufficient proof of the genuine physical basis of the afterglow streaks which, while not being explained finally in this paper (see below), assume a greater physical interest than they would have possessed if explained in the manner suggested by Flowers and others.

The use of pairs of photographs, such as those of Figs. 5–7, enable the true length, in three dimensions, of the spark channels to be found. This may be useful in experiments on the voltage drop in spark discharges, where the true channel length if it differs appreciably from the two dimensional value, should be known. This effect would probably be noticeable only with hydrogen channels; further experiments on these lines are in progress.

Before describing further analyses of records similar to those of Figs. 5 and 6, it is necessary to mention the very frequent and often complex division of channels noticed in argon and demonstrated in Figs. 6 (c), (d), and (e). The other gases so far examined show relatively infrequent branching which generally extends (Figs. 5 (a) and (d)) only for a fraction of the gap length. Argon, on the other hand, often shows (Figs. 6 (c) and (d)) complete branching of the gap by divided channels and also multiple division, a complex case of which is shown in Fig. 6 (e). The value of bi-directional photography in the elucidation of such phenomena is obvious. Yet another interesting effect is seen in Fig. 6 (c) (and many other records) where the afterglow streaks are seen to change their positions in space, as time advances. It seems probable in Fig. 6 (c) that the right-hand channel of the upper photo-

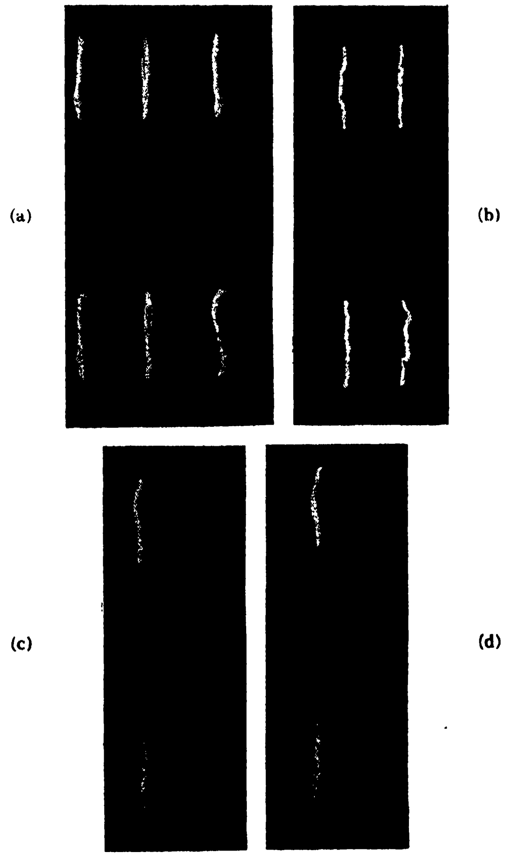


FIG. 5. Photographs of sparks in various gases; (a) oxygen, (b) hydrogen, (c) and (d) nitrogen.

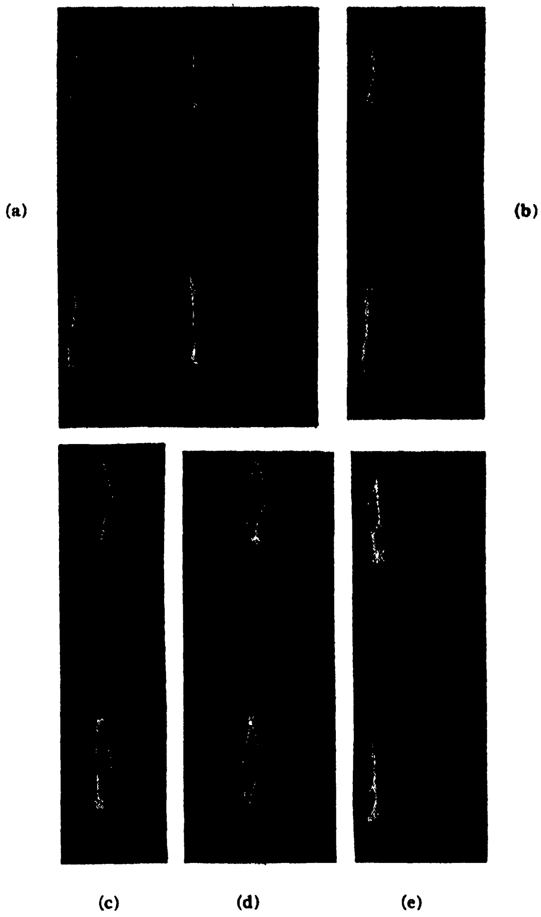


FIG. 6. Photographs of spark discharges in argon.

TABLE I.

| Gas | Total No. of sparks examined | No. of sparks with two channels | No. of sparks with three or more channels |
|---|---|---|---|
| Hydrogen | 27 | 0 | 0 |
| Oxygen | 43 | 27 | 5 |
| Nitrogen | 39 | 22 | 1 |
| Argon | 85 | 35 | 6 |
| Air | 71 | 10 | 0 |
| Methane | 20 | 1 | 0 |

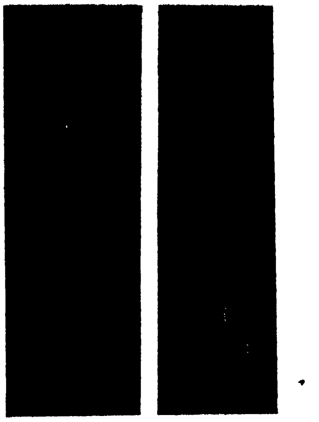


FIG. 7. Photographs of spark discharges in methane

graph is the left-hand channel of the lower photograph, since this branch shows, in both photographs, relatively little afterglow. These effects are being investigated further.

Table I gives some statistical data on channel division, which are discussed below.

The methane photographs (Fig. 7) were taken in the hope that the known high absorption of photo-ionizing radiations in such a complex molecular gas[18] would lead to the demonstration of unusual properties of the spark channel. The peculiar rippling structure of the methane channels are indeed noticeable, particularly on the leading edge of the channels where the afterglows, being absent, do not cause confusion. Such tortuous channels are to be expected, on general grounds, if the photo-ionization process at the streamer tip is inefficient.[1]

Microphotometer tracings of typical records in several gases were taken at right angles to the mean channel direction along afterglow streaks and across parts of the channel where such streaks were absent Figures 8 (a)–(c) show, respectively, tracings for argon, oxygen, and nitrogen. The right-hand tracings for each gas were taken along an afterglow. Figures 9 (a) and (b) show hydrogen records, the former taken in the afterglow. 1 cm on the time axis (right to

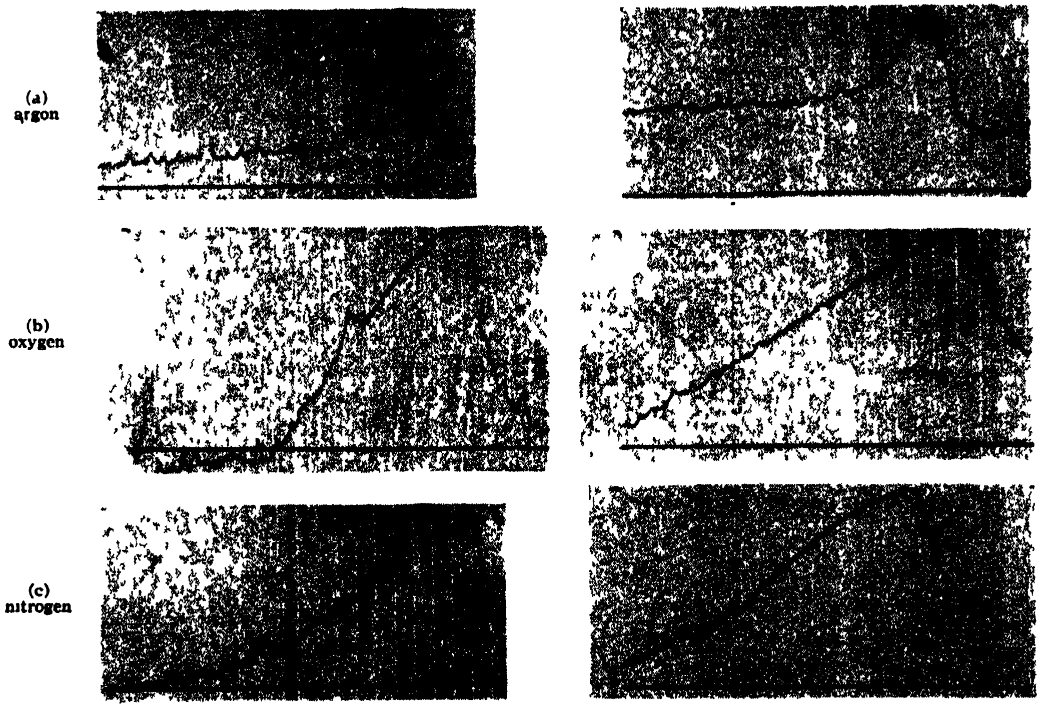


FIG. 8. Microphotometer tracings of spark photographs, such as those of Figs. 5–7. Time scale is from right to left. The left-hand records are for normal parts of the channels, and the right-hand records are for points in the channels which show the localized afterglows.

---

[18] S. A. Korff and R. D. Present, Phys. Rev. 65, 274 (1944).

left) of the figures represents 21 microseconds. It is possible to deduce from these records the light emission/time relation for any part of the channel if the response curve of the film is known. The latter quantity was determined by exposing samples of the film to 2000-amp. sparks, of the type studied in this paper, through an Ilford stepped filter (neutral) which was in turn calibrated with the usual photo-cell arrangement. Exposures were also made to steady light. A typical calibration record, using spark light, is shown in Fig. 10. Figures 11 and 12 show the results of the analysis, i.e., the gamma-curves for the film using spark and steady light, respectively.

Since the light from the spark channels is relatively very intense, with respect to the afterglows, the calibration records, such as Fig. 10, show the film's response to the channel and not the afterglow light. However, since Fig. 11 shows that the difference in the responses to channel and steady light is negligible and the response to the afterglow light must lie between the limits set by Fig. 11, it is reasonable to take the spark-light data of Fig. 11 as referring to the afterglow light.

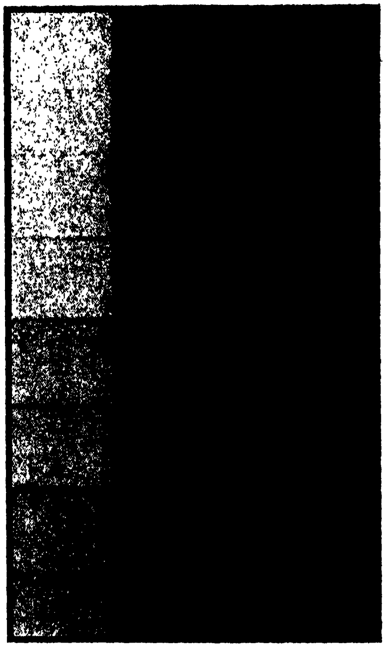


FIG. 10. Step wedge photograph with argon spark light, for film calibration purposes.

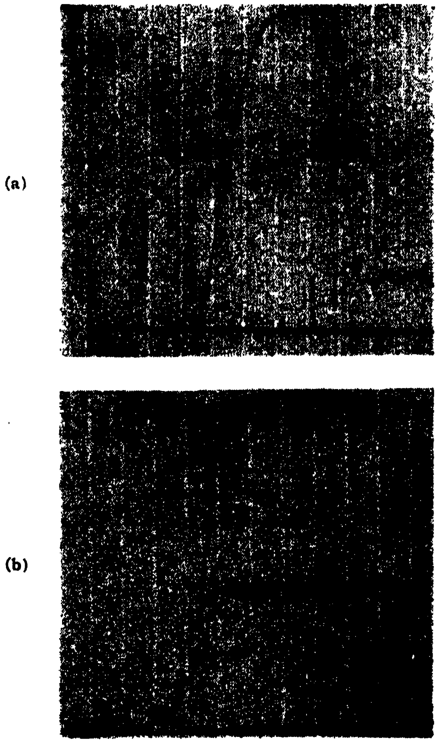


FIG. 9. Microphotometer tracings of hydrogen spark photographs (see Fig. (5b)); time scale is from right to left.

Figure 12 shows the final light intensity/time curve for an argon afterglow, with traces taken along an afterglow streak and across the channel at a point where no appreciable streak was observed.

## V. DISCUSSION OF THE RESULTS

It seems clear from the data given in Section IV that the suggestion of the bends being responsible, in a geometrical sense, for the afterglow streaks is untenable. The influence of the magnetic field of the discharge channel on the afterglow should be considered. This effect was suggested originally by Anderson and Smith[19] to account for similar localized afterglows, described as being due to vapor jets, with exploded wires, and it was pointed out that the magnetic field tended to prevent charged particles from leaving the conducting channel especially at the points of strongest fields. The field is weakest, however, at the outside of a sharp bend where afterglow

---

[19] J. A. Anderson and S. Smith, Astrophys. J. 64, 295 (1926).

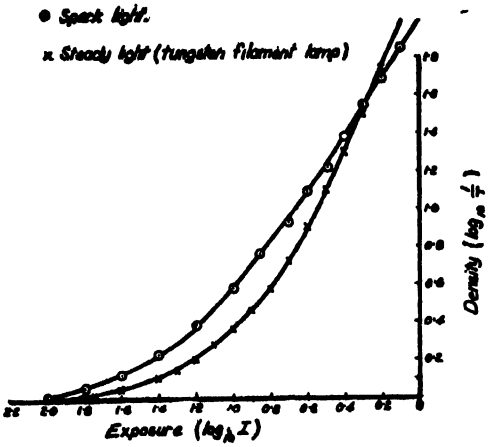


FIG. 11. Calibration of the photographic films with the step wedge.

streaks are often, but not always, encountered; and it is clear that no definite conclusion as to the relevance of this effect which could, at least, explain the localization of the afterglow can be made in the absence of further data.

The following mechanisms could account for general, i.e., non-localized afterglows. The influence of the low current still passing through the discharge channel after say 100 microseconds is not so easily dismissed. Figure 4, which is a reproduction of a current oscillogram for argon (no apparent systematic difference in residual currents for the different gases was noticed, but the accuracy of estimation of these low currents by the present method is not great) shows that the current at 100 microseconds from the beginning of current flow is about 20 amp. The records showed currents varying from this order of magnitude to almost zero. However, the passage of a small residual current cannot by itself account for the localized nature of the afterglows, since the current must also flow through the relatively dark regions of the discharge. It has been shown[4] that the afterglows persist even when square-wave pulses of current are used, and it therefore seems reasonable to postulate the existence of afterglows more or less independent of even a small current flow in the gap.

Afterglows in low pressure discharges are well known, by the pioneer work of Rayleigh, and later, many others on active nitrogen. Kenty[20] has observed an afterglow in argon, following a low pressure arc discharge, and has attributed this to electron/ion recombination. Afterglows in luminous clouds, following arc discharges at atmospheric pressure, have been noticed by O'Doherty[21] and others.

The afterglows in hydrogen have been discussed elsewhere, and it is sufficient here to state that they are probably due to electron/ion recombination. The afterglows in argon (and certain other gases, e.g., oxygen and nitrogen) could presumably be due either to recombination of various kinds (electron/ion mechanism in argon, where attachment is small) or to the persistence of metastable atoms. It is hoped that further experiments by the authors will clarify the matter.

It seems unlikely that persistence of metastable atoms in argon is the sole cause of the afterglows, since addition of oxygen or hydrogen to argon in reasonable quantities has only a minor effect on them,[4] although both diluents in very small amounts have a great effect on the dielectric recovery of argon spark gaps. If, therefore, electron/ion recombination is the main cause of the visible afterglow streaks it is of interest to consider why the afterglows in hydrogen and argon should be widely different in duration. The recombination coefficients themselves are not likely to be widely different, and the problem would then resolve into a study of the energy losses (by elastic collisions) of the electrons in the afterglow, since the recombination cross sections increase with decreasing electron energy. It is apparent at once from the known data on the Ramsauer effect etc., that for equal electron energy, the elastic collision cross sections are much greater in hydrogen than in argon for low electron energies, so that the electrons would attain thermal energies more quickly in the former gas and would therefore be removed by recombination more quickly. Unfortunately, Ramsauer data are not available for electron temperatures of ∼500–1000° (mean energy ∼0.1 volt or less), which probably obtain in the afterglow conditions.[22] Cravath[23] has worked out the average energy loss ($f$) per encounter, in elastic

---

[20] C. Kenty, Phys. Rev. 32, 624 (1928).

[21] J. J. O'Doherty, Nature 153, 558 (1944).

[22] J. D. Craggs and W. Hopwood, in the press.

[23] A. M. Cravath, Phys. Rev. 36, 248 (1930).

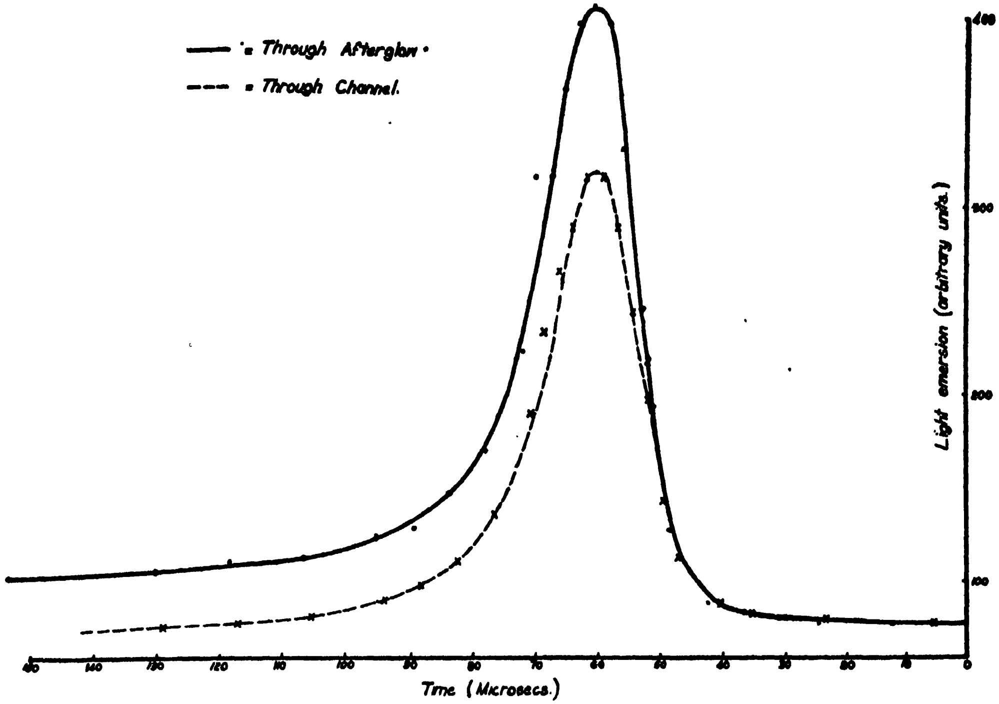


FIG. 12.

collision, expressed as a fraction of the average energy of the electron:

$$f=\frac{8}{3}\frac{mM}{(m+M)^2}\left(1-\frac{T_M}{T_e}\right),$$

where $T_M$ is the equivalent temperature of the molecules present and $T_e$ is the equivalent temperature of the electrons. Generally, $T_M \ll T_e$ so that

$$f \simeq \frac{8}{3}\frac{mM}{(m+M)^2}$$

($m+M$ are the electronic and molecular masses, respectively). It can be shown that $f$ is very great until $T_e \sim T_M$. Using elastic cross sections ($P_e$) given by Normand[24] and Brode,[25] the data of Table II are easily calculable. In some cases (at the lowest temperatures) extrapolation of the elastic cross-section/electron energy curves has been necessary; this is clearly a most unsatisfactory procedure, but it is the only possible one in the circumstances and leads only to approximate results. $T_M$ has been taken at 300°K. If a higher value for $T_M$ is permissible, then $f$ will be lower than the respective values in Table II, particularly at $T_e \sim 500$°K. The procedure for determining the total $f$ is (i) calculate the mean electronic energy and velocity ($v$) from $T_e$, which is assumed to correspond to a Maxwellian distribution; (ii) determine the mean free path $\lambda$ for elastic collisions, i.e., $1/P_e$, for the known electronic energy; (iii) from $\lambda$ and $v$ the total number of collisions in a given time, for each one of which a fractional energy loss $f$ is suffered, can be found, and so (iv) the total $f$ for that time can be computed.

TABLE II.

| Gas | Electron temp. °K | Total $f$ for time of $10^{-8}$ sec. |
|---|---|---|
| Hydrogen | 10,000 | 32 |
| Hydrogen | 1,000 | 39 |
| Hydrogen | 500 | 20 |
| Argon | 10,000 | 0.093 |
| | 5,000 | 0.066 |
| | 1,000 | 0.062 |

It must be emphasized again that the results are only approximate, but they show that the energy loss for electrons in hydrogen in our par-

---

[24] C. E. Normand, Phys. Rev. 35, 1217 (1930).
[25] R. B. Brode, Rev. Mod. Phys. 5, 257 (1933).

ticular conditions is far higher than for argon, and so a recombination spectrum would persist for a longer time (other things being equal) in argon than in hydrogen. The important influence of $T_M$ at values of $T_e \sim T_M$, and our ignorance of its magnitude, render further speculation unprofitable. Experiments now being made will, it is hoped, provide more accurate data on the values of $T_M$ and $T_e$ obtaining in the decaying channels of spark discharges.

## VI. CONCLUSIONS

1. Photographs have been taken with a rotating film camera, and simultaneously in two perpendicular directions of spark discharges.

2. The photographs show, we believe conclusively, that the localized afterglow streaks seen in records of this kind are not due to bends, in the spark channels, causing artificially long exposures by virtue of the enhanced depth of glowing gas, etc.

3. The causes of the afterglows are not known with certainty in all cases.

4. We believe that further experiments, now in hand, will give further information on localized afterglows in spark channels and probably, therefore, on ball or bead lightning[26] which we consider to be probably due to the same causes.

## VII. ACKNOWLEDGMENTS

The authors are grateful to Mr. F. R. Perry for the interest he has shown in, and the support he has given to, the work. They also wish to thank Mr. G. C. Williams and Mr. W. Dale for assistance with the camera and high speed oscillograph, respectively. Finally thanks are expressed to Sir Arthur P. M. Fleming, C.B.E., D.Eng., Director of Research and Education, and to Mr. B. G. Churcher, M.Sc., M.I.E.E., Manager of Research Department, Metropolitan-Vickers Electrical Company, Ltd., for permission to publish this paper.

---

[26] D. G. Beadle, Nature 137, 112 (1936).

---

# Erratum: Conductivity of Metallic Surfaces at Microwave Frequencies

[J. App. Phys. 18, 629 (1947)]

E. MAXWELL
*Massachusetts Institute of Technology, Boston, Massachusetts*

1. p. 629, line 5 of fourth paragraph: Replace $J_0^2/2\sigma\delta$ by $(J_0/2)^2(\delta/\sigma)$.

2. p. 629, line 8 of fourth paragraph: Replace $J_0$ by $J_0/\sqrt{2}$.

3. p. 629, second column, equation should read:

$$\int\int\left(\frac{J_0}{2}\right)^2\frac{\delta}{\sigma}ds=\int\int\frac{H_t^2}{2\sigma\delta}ds.$$

4. p. 632, Fig. 3b should be inverted.

5. p. 634, line 21 should be changed to read "*a* and *c* are incommensurable."

6. p. 636, line 25, replace "conductivity" by "resistivity."

# Letters to the Editor

## Quantitative Analysis of Mixed Powders with the Geiger-Counter X-Ray Spectrometer

ZIGMOND W. WILCHINSKY
*Esso Laboratories, Standard Oil Company of New Jersey, Baton Rouge, Louisiana*

July 23, 1947

A USEFUL characteristic of a Geiger-counter spectrometer is its ability to measure diffracted x-ray intensities for several samples in succession, on the same intensity scale. This property greatly facilitates the quantitative determination of crystalline components in a sample of mixed powders. In particular, it becomes quite easy to determine the amount of a crystalline component in the presence of other components which may be crystalline or amorphous, identified or unidentified. A simple expression relating the weight fraction of a crystalline component to experimental quantities is developed below.

For a flat sample used in a North American Philips x-ray spectrometer,[1] the relationship between the sample and the incident and diffracted x-ray beams is indicated in Fig. 1. Let us assume that simple x-ray diffraction theory is applicable. Then, for a chosen diffraction line of the *i*th crystalline component, the contribution to the integrated intensity from the elemental layer $dx$ is

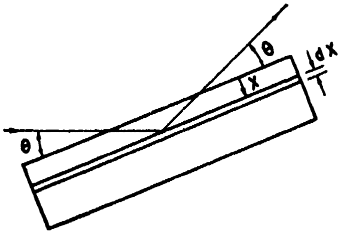


FIG. 1. Diffraction of x-rays from a flat sample in a Geiger-counter x-ray spectrometer.

$$dP_i=\frac{AJ_i}{\sin\theta}W_i\rho\exp(-2\mu_l x/\sin\theta)dx, \tag{1}$$

where $A$ = effective cross-section area of the beam at the sample, $J_i$ = contribution to the integrated intensity per unit weight of sample on the basis of no absorption, $W_i$ = weight fraction of the *i*th component in the sample, $\rho$ = mean density of the sample, $\mu_l$ = linear absorption coefficient for the sample. For a sample of sufficient thickness to be considered opaque, $P_i$ is obtained by integrating the right-hand side of (1) from $x=0$ to $x=\infty$, the result being

$$P_i=\tfrac{1}{2}AJ_iW_i\rho/\mu_l=\tfrac{1}{2}AJ_iW_i/\mu, \tag{2}$$

(2) where $\mu=\mu_l/\rho$ is the mass absorption coefficient for the sample. The factor $AJ_i$ can be evaluated by considering a sample consisting only of the *i*th component; the integrated intensity, $P_{0i}$, for this case, by Eq. (2) is

$$P_{0i}=\tfrac{1}{2}AJ_i/\mu_i, \tag{3}$$

where $\mu_i$ is the mass absorption coefficient for the *i*th component. Eliminating $AJ_i$ from (2), one obtains the useful and very simple expression:

$$W_i=\frac{P_i}{P_{0i}}\frac{\mu}{\mu_i}. \tag{4}$$

The value of $\mu_i$ can be readily computed from the chemical formula of the *i*th component and a table of mass absorption coefficients for the elements. Similarly, the value of $\mu$ can be computed from an elemental analysis for the sample.

---

[1] H. Friedman, Electronics 18, 132 (1945).

---

## The Conversion Factor for *kX* Units to Angström Units

ELIZABETH ARMSTRONG WOOD
*Bell Telephone Laboratories, Murray Hill, New Jersey*

July 21, 1947

X-RAY wave-lengths have been expressed in *X* units. The *X* unit is defined in terms of the calcite spacing and is nearly $10^{-11}$ cm, but is now known to differ from $10^{-11}$ cm by about 0.2 percent. During the last twenty-five years x-ray diffraction workers have expressed x-ray wave-lengths and crystal dimensions in terms of a unit which was 1000 *X* units, but instead of calling it 1000 *X* units have erroneously called it an Angström unit. In recent years, the x-ray diffraction groups have agreed to use the term kilo *X* unit (abbreviated *kX*) in place of the incorrectly used Angström unit, until agreement was reached on the best conversion factor to use for converting from *kX* to Angström units. Agreement on the factor has now been reached.

As Secretary of the American Society for X-Ray and Electron Diffraction, I have been instructed to call the attention of American x-ray workers to the following announcement which appeared in the January 1947, issue of the Journal of Scientific Instruments. Because of its importance it is here reproduced in its entirety.

At the annual conference of the X-ray Analysis Group of the Institute of Physics in July 1946 it was announced that agreement had been reached concerning the factor for converting measurements in *kX* units to Angström units. The factor agreed upon, after consultation with the American Society for X-ray and Electron Diffraction and Prof. Siegbahn was 1.00202. This factor is probably correct to 0.003 percent. Since wave-lengths in X-units have been measured to an accuracy of 0.001 percent, the wave-lengths in Angström units can be taken as accurate to 0.004 percent in general.

The following is a list of values of wave-lengths in Angström units of certain emission lines and absorption edges in common use. The column headed $K\alpha$ gives the mean value of $K\alpha_1$ and $K\alpha_2$, $K\alpha_1$ being allowed twice the weight of $K\alpha_2$.

Current values of the physical constants, such as those quoted by Birge in the 1941 volume of the Physical Society's *Reports on Progress in Physics*, should be used in conjunction with these wave-lengths. In particular density $\rho$ is given by the equation

$$\rho = 1.66020 \Sigma A / V$$

where $\Sigma A$ is the sum of the atomic weights of the atoms in the unit cell, and $V$ is the volume of the unit cell in $A^3$.

|  | $K\alpha_1$ | $K\alpha_2$ | $K\alpha$ | $K\beta_1$ | Absorption edge |
|---|---|---|---|---|---|
| Cr | 2.28962 | 2.29352 | 2.2909 | 2.08479 | 2.0701 |
| Mn | 2.10174 | 2.10570 | 2.1031 | 1.91016 | 1.8954 |
| Fe | 1.93597 | 1.93991 | 1.9373 | 1.75654 | 1.7429 |
| Co | 1.78890 | 1.79279 | 1.7902 | 1.62073 | 1.6072 |
| Ni | 1.65783 | 1.66168 | 1.6591 | 1.50008 | 1.4869 |
| Cu | 1.54050 | 1.54434 | 1.5418 | 1.39217 | 1.3802 |
| Zn | 1.43510 | 1.43894 | 1.4364 | 1.29520 | 1.2831 |
| Mo | 0.70926 | 0.71354 | 0.7107 | 0.63225 | 0.6197 |
| Rh | 0.61326 | 0.61762 | 0.6147 | 0.54559 | 0.5341 |
| Pd | 0.58545 | 0.58982 | 0.5869 | 0.52052 | 0.5090 |
| Ag | 0.55941 | 0.56381 | 0.5609 | 0.49701 | 0.4855 |

It is recommended that in any published work the values of the wave-lengths used should be explicitly stated.

W. L. BRAGG
Chairman, X-ray Analysis Group of the
Institute of Physics (England).

# New Books

## Astronomy

By JOHN C. DUNCAN. Fourth Edition. Pp. 500. Harper and Brothers, New York, 1946.

Professor Duncan's book now in its fourth edition offers an excellent text for the beginning college student of astronomy. This book may be read with profit by anyone who wishes to obtain a knowledge of modern astronomy.

The introductory chapter includes an exceptionally fine set of star charts showing the naked eye stars, though a more detailed description of how to use them would have helped the student. Following the chapter on the celestial sphere and coordinate systems is one on telescopes. Though the Schmidt telescope is described, the reader is not given the simple explanation of the function of the correcting plate. A new and very excellent chapter on Celestial Navigation has been added following the chapter on the Earth and Sky. Chapters on the Orbital Motion of the Earth, the Moon, and Eclipses are followed by a chapter on Spectroscopy which is introduced early in order to make use of the principals involved in the chapter on the Sun which follows. The Paths of the Planets and the Law of Gravitation are next taken in order; the latter seems rather detailed in parts for a book designed for beginning students. Chapters on the Planets, Meteors, and Comets precede a discussion of the evolution of the solar system, a chapter which dwells unduly long on the older theories at the expense of the more modern theories of Spitzer, Littleton, and Russell. However, since even modern theories on this subject are still far from giving us a complete answer to this riddle, perhaps Dr. Duncan is justified in these omissions.

The stars and their motions are next discussed and then follows a completely up-to-the-minute (1945) chapter on Physics of the Stars. Here the student may find "the going a little tough," and will undoubtedly want to review the earlier chapter on Spectroscopy before wrestling with even elementary astrophysics. After a chapter on Variable Stars is one on Star Clusters and Nebulae in which the careful student may be somewhat confused to find that the Magellanic Clouds are discussed as star clouds in our own galaxy rather than as external galaxies which are satelites of our system. The concluding chapters on the Galactic System and Beyond the Milky Way are excellently treated. One might desire a more detailed discussion of the age of the universe which indicates that we are relatively young ($10^9$ or $10^{10}$ years old) and not a decrepit $10^{12}$ years as Professor Duncan states.

The appendix to the book includes many useful tables which in the earlier editions had been scattered through the book. A table for obtaining the Julian Day for any date from 1940–1990 is an especially welcome addition.

Teachers will wish that the paragraphs were numbered for ease in making assignments, but this is more than compensated for by a large selection of excellent problems involving simple mathematical manipulation.

The book is excellently written, is thoroughly up to date, and includes a nice balance of historical background. It is by far the most attractive general book on astronomy ever written, and the illustrations, many of which were made by Dr. Duncan, are magnificent. The book will undoubtedly be even more popular than the widely used earlier editions, and rightly so.

CARL K. SEYFERT
*Vanderbilt University*

# New Booklets

The following recent publications of the National Bureau of Standards are of interest to physicists: RP 1790, *Infrared Emission Spectra of Krypton and Argon*, Curtis J. Humphreys and Earle K. Plyler, 10 cents; RP 1793, *Changes Caused in the Refractivity and Density of Glass by Annealing*, Arthur Q. Tool, Leroy W. Tilton, and James B. Saunders, 10 cents.

Beginning with the October issue, the quarterly journal *Mathematical Tables and Other Aids to Computation* will publish a new feature section, "Automatic Computing Machinery," designed to disseminate information and news on research and development in the field of high speed automatic calculating machinery. Contributions to this section are invited and should be addressed to Dr. E. W. Cannon, Head of the Mathematics Group, Machine Development Laboratory, National Bureau of Standards, Washington, D. C. Calendar year subscription to the journal is $4; single numbers cost $1.25. Payments should be made to the National Academy of Sciences.

*Tracerlog* for August 1947, published by Tracerlab, Inc., 55 Oliver Street, Boston 10, Massachusetts, describes the film badge service offered by Tracerlab for health monitoring purposes where employees are exposed to radiation. 16 pages.

Sound Apparatus Company, 233 Broadway, New York 7, New York, recently brought out two bulletins of interest to physicists: *Impedance Vectorgraph* (4 pages) and *Portable Reverberation Analyzer* (12 pages). Because of the highly specialized nature of the latter brochure, the Company is interested in sending it only to people actually engaged in acoustical measuring work.

North American Philips Company, 100 East 42 Street, New York 17, New York, has announced two publications that are available without charge: an 8-page booklet on quality control, and a 4-page folder (R1050) entitled *Geiger-Mueller Counters Applied to Mining.*

*The Frontier* for June 1947, published by Armour Research Foundation of Illinois Institute of Technology, contains the following featured articles: "The Application of Industrial Research," "Noise and Vibration Reduction," "The Mechanism of Cavitation Erosion," "Crystal Structure of Nicotinamide," and "Contact Angle Measurements in Lithographic Research."

# Here and There

### Personnel

The University of Wisconsin has announced the retirement in June of Florence E. Allen, Assistant Professor of Mathematics, and John R. Roebuck, Professor of Physics.

Two promotions in the Department of Physics at New York University were announced recently. Morton Hamermesh was elevated to an Associate Professorship and Serge A. Korff to a full Professorship.

Donald A. Quarles, Director of Apparatus Development of Bell Telephone Laboratories, was elected a Vice President at a meeting of the board of directors in July.

Kinichi Watanabe, formerly of the University of Hawaii, has been appointed Assistant Professor in the Department of Physics at Wabash College.

### College of Horology and Instrument Technology

A National College of Horology and Instrument Technology is being set up in the Northampton Polytechnic, London, England, to open in October of this year. Full information may be obtained by addressing Mr. R. A. Fell, Northampton Polytechnic, St. John Street, London, E.C. 1.

### Institute of Numerical Analysis Established

Plans have been completed for the establishment of one of the newest units of the National Bureau of Standards—the Institute of Numerical Analysis—at the University of California. One of the giant high speed electronic computing machines now under development by the Bureau of Standards will be installed at the Institute when completed. The Institute has two primary functions. The first is research in applied mathematics aimed at developing methods of analysis which will extend the use of the high speed electronic computers. The second is to act as a service group for western industries, research institutions, and government agencies. The service function will include not only the use of the machines for problem solving but also assistance in the formulation of problems in applied mathematics of the more complex and novel types.

### Induction Heating Corporation

The Induction Heating Corporation, manufacturer of induction and dielectric heating equipment, has moved to new quarters at 181 Wythe Avenue, Brooklyn 11, New York.

### Fellowship on Waterproofed Paper

A comprehensive investigation of waterproofed papers, used for such purposes as packaging, curing of concrete, and moisture barriers in buildings, is being undertaken at the National Bureau of Standards through a fellowship of the Waterproof Paper Manufacturers Association. A long-range research program, extending over at least three years, has been jointly agreed upon. The services of Robert C. Reichel as Research Associate are being provided by the Association, with laboratory facilities and technical assistance furnished by the Bureau.

### Applied Mathematics Symposium

The first annual symposium in applied mathematics of the American Mathematical Society was held at Brown University August 2–4. The subject of the symposium was "Non-Linear Problems in Mechanics of Continua." The American Institute of Physics, the American Society of Mechanical Engineers, and the Institute of the Aeronautical Sciences were co-sponsors.

### Contractor for Brookhaven Atomic Pile

The H. K. Ferguson Company will build the atomic pile and associated facilities at Brookhaven National Laboratory for atomic research on Long Island, New York. Construction has begun and will be completed in 1948.

### Honorary Degrees

The following scientists received honorary degrees recently: John H. Van Vleck, Harvard University, the D.Sc. from the University of Wisconsin; Roger Adams, University of Illinois, the D.Sc. from the University of Pennsylvania; Vannevar Bush, Carnegie Institution of Washington, the D.Sc. from Columbia University.

### New Course at Yale

Yale University has announced the inauguration of a new course to be offered this year on the Development of the Sciences. It is designed to aid in "the closer integration of the sciences and the humanities in the college curriculum and to emphasize the contribution of the sciences to liberal education." The course will cover two terms. Representing the science of physics is Henry Margenau.

### Pittsburgh Spectroscopy Conference

The Eighth Pittsburgh Conference on Applied Spectroscopy, sponsored by the Spectroscopy Society and the University of Pittsburgh, will be held Thursday to Saturday, November 13 to 15, 1947, inclusive, at the Mellon Institute Auditorium. As in the past, sessions on both emission and absorption spectroscopy are planned. Address inquiries to Mary E. Warga, University of Pittsburgh, Pittsburgh 13, Pennsylvania.

### Ionosphere Studies Truck Station

A major problem in the study of long distance radiowave propagation is the role of the ionosphere in returning radio waves of suitable frequency back to the earth's surface. This study may best be accomplished by pulse techniques as used in radar. In fact, the radar techniques developed during the war were based, in great part, on work of this nature previously developed in connection with ionosphere problems.

Up to the present time most ionosphere studies have been confined to the vertical incidence case for reasons of simplified instrumentation and theoretical analysis. The oblique incidence and practical case has been studied by remote station field intensity recording and by theoretical analysis. But little oblique incidence pulse work has been done up to the present time because of the relatively complex instrumentation required. This is the purpose of the equipment set up in the truck station illustrated on the front cover.

The truck contains complete facilities for recording manually or photographically the pulse emissions from a high power ionosphere transmitter located at the Pennsylvania State College. Moderately high power communications equipment is also included for communication with the college station over distances of several hundred miles. A power unit for supplying the equipment is contained in the trailer at the rear of the truck. The pulse equipment receiving antenna is supported by two sectionalized masts.

For operation, the truck is driven to the place where observations are to be made, the two antennas are erected and, by pushing a button, the power unit is turned on. The equipment is then ready for operation. Experience has shown that this may all be done in about five minutes. Hence, if desired, observations may be conducted at many distances or locations with respect to the transmitting position in a single day.

Several studies utilizing this equipment are currently in progress. This work is being done by Professor A. H. Waynick of the Electrical Engineering Department under contract between the Pennsylvania State College and the Watson Laboratories, U. S. Army Air Forces.

### Calendar of Meetings

**October**

| | |
|---|---|
| 15-18 | American Society of Civil Engineers, New Orleans, Louisiana |
| 20-24 | American Society for Metals, Chicago, Illinois |
| 21-23 | American Standards Association, New York, New York |
| 21 25 | American Chemical Society, California Section, San Francisco, California |
| 23 25 | Optical Society of America, Cincinnati, Ohio |
| 30-31 | National Research Corporation and the American Chemical Society (High Vacuum Symposium), Cambridge, Massachusetts |

**November**

| | |
|---|---|
| 3-5 | National Electronics Conference, Chicago, Illinois |
| 3-7 | American Institute of Electrical Engineers, Chicago, Illinois |
| 6-7 | Society of Automotive Engineers (Fuels and Lubricants Meeting), Tulsa, Oklahoma |
| 7-8 | Conference on X-Ray and Electron Diffraction, Mellon Institute of Industrial Research, Pittsburgh, Pennsylvania |
| 9-12 | American Institute of Chemical Engineers, Detroit, Michigan |
| 13-15 | Conference on Applied Spectroscopy, Mellon Institute, Pittsburgh, Pennsylvania |
| 17-19 | Institute of Radio Engineers and the Radio Manufacturers' Association, Engineering Department, Rochester, New York |
| 28-29 | American Physical Society, Houston, Texas |

**December**

| | |
|---|---|
| 1-3 | Society of Automotive Engineers (Air Transport Engineering Meeting), Kansas City, Missouri |
| 1-5 | American Society of Mechanical Engineers, New York, New York |
| 26-31 | American Association for the Advancement of Science, Chicago, Illinois |
| 29 31 | American Physical Society, Chicago, Illinois |

Journal
of
Applied Physics

Volume 18, Number 11 November, 1947

# Initial Oscillation on a Rotational Viscometer

RUTH N. WELTMANN
*Interchemical Corporation, New York, New York*

(Received May 7, 1947)

The oscillatory system of a rotational viscometer has been investigated. The oscillation equations were checked with experimental bob deflection-time curves obtained for Newtonian liquids.

To be able to rapidly obtain correct values when taking flow curves the oscillatory system should be critically damped. Since this condition cannot be easily obtained in practice, the system should be slightly overdamped, meaning that the inertia ($I$) should be equal or smaller than $\pi\mu^2/720KS^2$. It is shown that due to the higher viscosity of thixotropic materials at lower rates of shear, the flow curves at these rates of shear give an indication of rather too low than too high shearing stresses.

THE flow characteristics of thixotropic substances are known to change as a function of time, when these materials are subjected to shearing forces. However, there is little information available about the change in structure of the material during the initial period of agitation, since no instantaneous readings can be made with most systems of measurement. For studying the structural changes in the initial period, the use of recording equipment is most desirable since it provides a continuous graph of the position of the indicating means of the viscometer, thus seemingly allowing an interpretation of the flow behavior of the material from the start of motion.

This paper deals with the problems encountered with a rotational viscometer when changing the cup speed from one to another angular velocity. Details of the viscometer have been described by Green.[1] The recorder which has been used in conjunction with this viscometer has also been previously discussed.[2]

However, even the recorded values are not always true representations of the flow characteristics of the investigated material, since the indicating means which control the recording mechanism do not instantaneously adjust themselves to the proper position but require time to perform the necessary motion.

In a rotational viscometer the shearing force at any rate of shear is obtained from the angular deflection of a bob, but only after the bob has come to an equilibrium position. In Green's design of a viscometer the bob is suspended on a helical spring. The stiffness of the spring and the inertia of the bob structure form an oscillatory system in which damping is provided by the viscosity of the test material which is located between the cup and the bob. When the cup of the rotational viscometer is set into motion for the purpose of performing a viscometric measurement the bob starts to move from its zero position until it reaches its final deflection. The time which is required to complete the bob motion depends upon the three elements of the oscilla-

---

[1] H. Green, Ind. Eng. Chem. Anal. Ed. 14, 576 (1942).
[2] R. N. Weltmann, Rev. Sci. Inst. 16, 7, 184 (July, 1945).

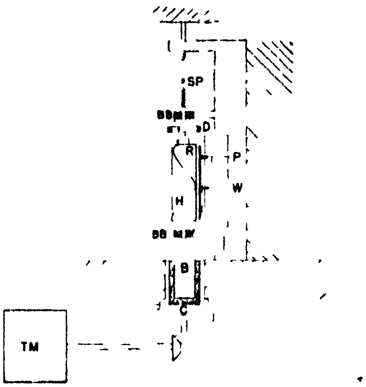


FIG. 1. Schematic viscometer and recorder assembly.

tory system, the stiffness of the spring, the inertia of the bob parts and the viscosity of the material. Though reference has been made to a condition which exists only when starting a viscometric measurement, the same reasoning also applies when the cup speed is changed step by step from one speed to another to obtain a rate of shear—shearing stress diagram.

It is of greatest importance to design properly the spring-bob combination so that the bob will come to its equilibrium condition in the shortest period of time. An analysis is being made of the most suitable physical dimensions of the elements of the oscillatory system for various test materials.

## INSTRUMENTATION

Figure 1 is a schematic sketch of the measuring and recording means of a rotational viscometer. The test material is contained between cup (*C*) and bob (*B*). The cup which is placed in a 30°C constant temperature bath is driven at various speeds to obtain different shearing forces in the material. The bob, supported between two ball nearings (*BB*), deflects until the torsional force of the spring (*SP*) equals the resistance force in the test material. This deflection is a measure of the applied shearing stress. The bob carries the deflection indicating disk (*D*) for visual measurements and the movable registering elements of the recorder (*R*). The recorder consists of two electrodes between which a high frequency spark is generated. The one electrode is a straight wedge (*W*), while the other electrode is a helical wire (*H*) wound around a cylinder concentric to the bob shaft, so that each position of the wire with respect to the straight wedge coincides with a unique deflection reading. The recording paper (*P*) travels between these electrodes at a constant speed, being moved by a slow constant speed motor (2.9 cm/sec.), and holes are pierced in the paper whenever a spark jumps over the gap. Thus the continuation of pierced points on the paper represents a bob deflection-time curve. On recording a flow curve the paper is coupled to the drive mechanism, the transmission motor combination (*TM*), in such a manner that the paper position becomes a function of the cup speed.

## CUP INERTIA

The cup as part of the drive mechanism is expected to obtain a constant angular rotation which is being impressed on it by the synchronous motor and transmission combination as soon as the initial acceleration period is over. This has been checked experimentally by rigidly connecting the cup with the bob from which temporarily the spring was removed. Wax was poured in the clearance between cup and bob and left to solidify before the test. The deflection of the bob as a function of time was recorded. Small curvatures are obtained at the beginning of the straight lines characteristic for each applied speed, as shown in Fig. 2. This is evidence that the cup needs a certain time before it attains its given speed. This time lag of the cup was determined by extending the straight line to zero deflection. From numerous experiments performed at various speeds this time intercept was found to average about 0.14 second. For further reference it will be designated as $t_c$.

## OSCILLATORY SYSTEM

The force equilibrium of the oscillatory system in such a rotational viscometer is given by the basic differential equation,

$$I\frac{d\alpha^2}{dt^2}+\frac{\mu}{S}\frac{d\alpha}{dt}+K\frac{360}{2\pi}\alpha=\frac{\mu}{S}\omega, \qquad (1)$$

$I$—the inertia of the bob, bob shaft, and deflection indicating dial in gram $\text{cm}^2$,
$\alpha$—the deflection in radians,
$\mu$—the viscosity in poises,

$t$—the time of motion in seconds,
$S$—an end effect corrected instrumental constant $S=[(1/R_b^2-1/R_c^2)/4\pi(h+h_e)]$,
$R_c$—radius of cup in cm,
$R_b$—radius of bob in cm,
$h$—immersed height of bob in cm,
$h_e$—hypothetical height added to bob to correct for end effect,
$K$—torsion constant of spring in dynes cm determined per degree deflection,
$\omega$—applied angular velocity in radians per second.

The end effect is the shear produced between the bottom of the cup and the bob. It can be experimentally obtained by measuring shearing stresses for various bob heights under otherwise constant conditions. Upon extrapolation of the shearing stress—bob height curve, which is a straight line, it is found that for zero shearing stress a negative bob height is required. This is the hypothetical bob height $h_e$—contributing a shear equal to the shear effect between the bottom of the cup and bob—which has to be added to the immersed bob height in the equation for the end effect corrected instrumental constant $S$. In the cup and bob assembly employed, the end effect amounts to about 2 percent and the corrected $S$ equals $2.24\times10^{-3}$.

Equation (1) is only strictly correct for materials whose viscosity is independent of time and rate of shear. It has three solutions:

(1) The overdamped or aperiodic oscillation, where the bob deflection increases slowly until the bob reaches its final position. If the bob creeps very slowly towards its maximum deflection there is some danger that the cup speed is increased to the next step before the bob has reached its final position and thus incorrect results will be obtained. Case 1 is realized when:

$$\frac{\mu^2}{4I^2S^2}>\frac{360K}{2\pi I} \quad \text{or} \quad I<\frac{\pi\mu^2}{720KS^2}. \tag{2}$$

Then the solution of Eq. (1), if $D$, the deflection in degrees, is substituted for $(360/2\pi)\alpha$, is:

$$D=\frac{\mu\omega}{2KS\Gamma}e^{-(\mu/2IS)t}[(-\mu/2IS-\Gamma)e^{\Gamma+} \\ -(-\mu/2IS+\Gamma)e^{-\Gamma t}]+\frac{\mu\omega}{KS},$$

where

$$\Gamma=(\mu^2/4I^2S^2-360K/2\pi I)^{\frac{1}{2}}. \tag{3}$$

(2) The underdamped or periodic oscillation, where the bob deflection increases rather fast but exceeds its final position and oscillates around this position before it ultimately reaches it. In this case the most erroneous values can be obtained giving either too low or too high readings depending on the time after which they are taken. Case 2 is realized when:

$$\frac{\mu^2}{4I^2S^2}<\frac{360K}{2\pi I} \quad \text{or} \quad I>\frac{\pi\mu^2}{720KS^2}. \tag{4}$$

With this condition and again substituting $D$ for $(360/2\pi)\alpha$, Eqs. (1) and (2) become:

$$D=-\frac{\mu\omega}{KS}e^{-(\mu/2IS)t}\left[\cos\Gamma t+\frac{\mu}{2IS\Gamma}\sin\Gamma t\right]+\frac{\mu\omega}{KS}$$

where

$$\Gamma=(360K/2\pi I-\mu^2/4I^2S^2)^{\frac{1}{2}}. \tag{5}$$

(3) The oscillation with critical damping where the bob reaches its final position fastest. This most desired case 3 is realized when:

$$\frac{\mu^2}{4I^2S^2}=\frac{360K}{2\pi I} \quad \text{or} \quad I=\frac{\pi\mu^2}{720KS^2}. \tag{6}$$

Then

$$D=-\frac{\mu\omega}{KS}e^{-(\mu/2IS)t}\left(1+\frac{\mu t}{2IS}\right)+\frac{\mu\omega}{KS}. \tag{7}$$

It is also interesting to investigate the case in which the inertia is negligibly small. Under these conditions when $I=0$, the first term of Eq. (1) is omitted and then the equation becomes:

$$\frac{\mu}{S}\frac{d\alpha}{dt}+K\frac{360}{2\pi}\alpha=\frac{\mu}{S}\omega. \tag{8}$$

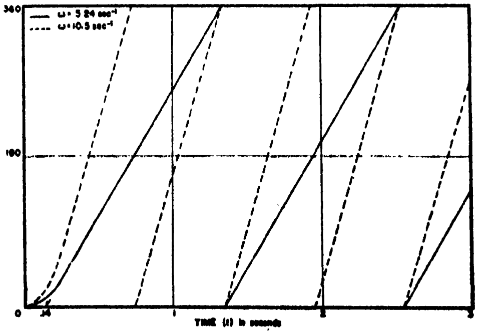


FIG. 2. Recorded cup deflections versus time of rotation for two different cup speeds to indicate cup inertia.

The solution of Eq. (8), if $D$ is again substituted for $(360/2\pi)\alpha$ is:

$$D=-\frac{\mu\omega}{KS}e^{-(360KS/2\pi\mu)t}+\frac{\mu\omega}{KS}. \tag{9}$$

Equation (7) represents the case in which the inertia in the oscillatory system has the highest permissible value without allowing the system to fall into a periodic oscillatory motion. Equation (9) on the other hand represents due to the removal of inertia the extreme case of aperiodic damping provided that everything else is maintained equal (same stiffness and same damping).

Equations (3), (5), (7), and (9) give the deflection ($D$) at any instant only if the angular velocity prevailing in the system before a change took place was zero. Otherwise $D$ is the deflection which has to be added to the prevailing deflection to obtain the new deflection value. The prevailing deflection is equal to $\mu\Omega/KS$, where $\Omega$ is the prevailing angular velocity.

## EXPERIMENTS

Experimental bob deflection-time curves for various rates of shear, test oils and bob inertias, were obtained with the recorder. As mentioned before, Eq. (1) is only correct when dealing with materials whose viscosity is independent of rate of shear and time of application at a constant rate of shear. Therefore, only true Newtonian materials can be used to check experimentally the solutions of Eq. (1). Oils, as shown in the literature,[3] are only true liquids up to their limiting rates of shear. Hence for these experiments they were tested at various rates of shear but only below the limiting rate of shear, where they are definitely known to behave like true liquids.

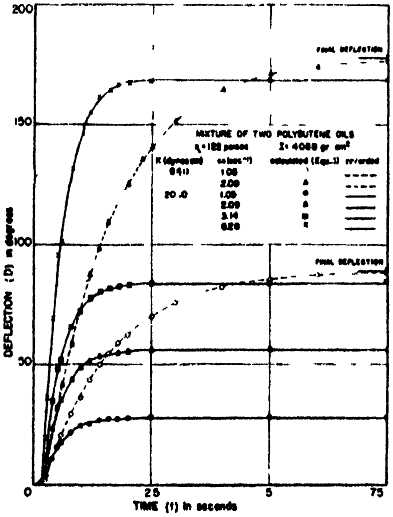


Fig. 3. Aperiodic oscillation curves for an oil in the Newtonian region, showing calculated points superimposed on experimentally recorded curves.

By using materials which varied in viscosity the damping of the oscillatory system was changed. In most cases for materials of different viscosity a change in the stiffness of the spring was required, since otherwise the deflection readings would have become either too small, thus lacking in accuracy, or too large so that the spring would have been overstrained.

In Fig. 3, bob deflection-time curves following Eq. (3), representing the case of over damped or aperiodic oscillation, are shown, for different applied rates of shear. Two bob springs of different stiffness were used. The same is shown in Fig. 4, but for an oil which gives underdamped bob deflection-time curves representing the case of periodic oscillation. In each case the cup started motion at zero time. Thus the prevailing angular velocity ($\Omega$) in both cases was zero. It was found that the equations check the experimental curves only when the time lag ($t_c$) of the cup is taken into consideration. This was done by deducting $t_c$ in each case from the measured time $t$. In Figs. 3 and 4 points obtained from Eqs. (3) and (5), have been super-imposed on the respective experimental curves with very good agreement.

In all cases so far described the inertia of the bob system was kept constant, ($I=4088$ g cm$^2$). In Fig. 5 curves are shown, representing conditions in which the inertia of the bob parts was varied in order to demonstrate that an increase in inertia can change the oscillatory system in such a manner that aperiodic oscillations will become periodic ones even though the same material and the same spring are used. Experimentally the inertia of the bob system was changed by operating without and with the deflection-indicating dial and by substituting the

---

[3] Weltmann, R. N., Ind. Eng. Chem. Anal. Ed. 15, 424 (July, 1943).

deflection-indicating dial with various lead disks. The weight of the lead disks was limited only by the carrying capacity of the spring. Thus the inertia was changed in various steps from 452 to 29308 grams cm². The bob deflection-time curves in Fig. 5 have been taken with one oil and at one applied rate of shear but with bob systems of different inertia. As can be seen two of these curves follow Eq. (3), while the other three follow Eq. (5). Again experimentally obtained and calculated bob deflection values are plotted.

### INERTIA OF THE TEST MATERIAL

Since the test material has to be accelerated when changing the cup speed, it was thought that the inertia of the test material in the viscometer might have to be taken into consideration in deciding whether under critical conditions periodic or aperiodic oscillation will result. To obtain experimental evidence about the inertia effect of the material the time-deflection curves of two substances of almost the same viscosity but of quite different specific gravity were recorded at various rates of shear. To exclude the stiffness element from the measurement the bob was decoupled from the spring. The time-deflection curves which were thus obtained were identical for the two test materials used which had a viscosity of about 5.5 poises. One of the two substances was a mineral oil with a specific gravity of 0.93 and the other one was a mercury potassium iodide solution in glycerine with a specific gravity of 2.84. The results are not surprising since the inertia of the test material even with the high specific gravity does not exceed 15 percent of that of the lightest bob system. The specific gravity of the mercury salt solution has been chosen to be high enough so that it will hardly be exceeded by any other material to be tested in the viscometer.

### PRACTICAL CONSIDERATIONS FOR NEWTONIAN LIQUIDS

In the measurement of flow curves of materials of which the flow characteristics had not been previously determined, it is important that the rate of shear can be increased and decreased at a very fast pace. On the other hand precautions must be taken that the rate of increase and decrease is not so high that the bob is prevented from attaining its final deflection value for each rate of shear step. The fastest adjustment of the the bob takes place when the oscillatory system is critically damped and it is thus this condition which is most desirable to maintain while performing a flow curve measurement. If the system is underdamped, a condition which will produce periodic oscillations whenever the rate of shear is changed, flow curves could be recorded which would provide shearing stress values for the different rates of shear which are either lower or higher than the correct values. For this reason an overdamped oscillatory system should be used in preference to an underdamped one. In order to accomplish this for materials of low viscosity an extremely low bob inertia is required.

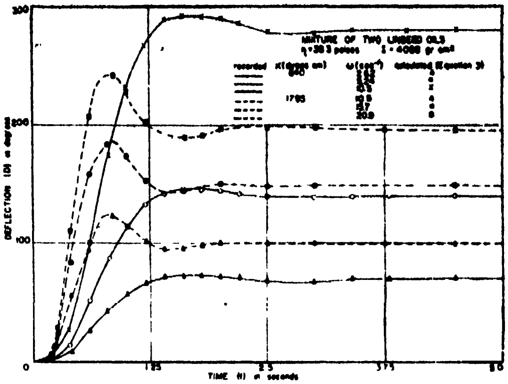


FIG. 4. Periodic oscillation curves for an oil in the Newtonian region, showing calculated points superimposed on experimentally recorded curves.

The bob reaches its final position after an infinite time has elapsed. This end value of the deflection is realized in all cases when,

$$D=\frac{\mu\omega}{KS} \quad \text{for} \quad \Omega=0 \qquad (10)$$

and

$$D=\frac{\mu(\Omega+\omega)}{KS} \quad \text{for} \quad \Omega\neq 0. \qquad (11)$$

Equation (10) represents not only the final deflection when the prevailing angular velocity was zero, but also equals the final deflection value which has to be added to any previously prevailing deflection.

Both Eqs. (10) and (11) will be recognized as the general viscosity equations for a rotational viscometer,[4] and are obtained from Eqs. (3),

---

[4] M. Reine and R. Riwlin, Kolloid Zeits. 43, 1 (1927).

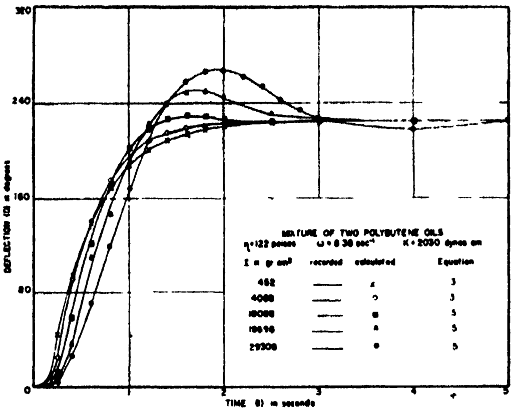


FIG. 5. Oscillation curves of an oil in the Newtonian region obtained with different bob inertias under otherwise constant conditions, showing calculated points superimposed on experimentally recorded curves.

(5), (7), and (9) for an infinite time. When $t$ is infinite the whole first term of all four equations becomes zero. However, for all practical purposes the end position can be assumed to have been reached when the deflection is about 1 percent less than the ultimate value. To obtain a value of 1 percent less than the deflection value $D$ of Eq. (10), the whole first term of Eqs. (3), (5), (7), and (9) has to be equal to the one percent which in turn means that the multiplicant of $\mu\omega/KS$ in the first term has to equal 0.01.

Since periodic oscillations are undesirable because of the possibility of obtaining not only too low, but also too high deflection readings and since the case of critical damping is difficult to achieve in practical tests, the damped oscillation curves will occur most frequently and therefore have been investigated further. It might be recalled that the two limiting conditions for the damped oscillation curves are the critical damping with the highest inertia before periodicity is obtained and the case where no inertia is present.

For the case of critical damping, which gives the fastest deflection movement, the equation which contains the time $t_F$ necessary to reach the final deflection value within 1 percent is,

$$e^{-(\mu/2IS)t_F}\left(1+\frac{\mu}{2IS}t_F\right)=0.01. \qquad (12)$$

By employing the graphical method,

$$t_F(\text{critical})\cong\frac{2IS\,6.64}{\mu} \qquad (13)$$

and since in this case Eq. (6) is valid,

$$t_t(\text{critical})\cong\frac{6.64\pi\mu}{360KS}. \qquad (14)$$

In the case that there is no inertia present the time for a deflection which is 1 percent short of its final position becomes from Eq. (9)

$$t_F(\text{inertialess})=-\frac{2\pi\mu\ln 0.01}{360KS}. \qquad (15)$$

Thus Eqs. (14) and (15) differ by a factor of about 1.4 That means, that everything else being equal, the fastest swing of the bob takes a little less than $\frac{3}{4}$ of the time which would be required for a swing with an inertialess system. Equation (15) can be written:

$$t_F=0.08\mu/KS=0.08D/\omega. \qquad (16)$$

In prior experimental work the author has taken most flow curves in one of the following manners. In the majority of cases the speed of the cup was raised in about 10 distinct steps during a period of about 26 seconds to cover a change of angular cup velocity of about 10 reciprocal seconds. For special runs the change of angular cup velocity of about 10 reciprocal seconds was attained in 13 seconds by taking only 5 steps. Under both these measuring conditions the time available for adjustment of the bob to its final position was 2.6 seconds. Thus $t_F$ in Eq. (16) should not exceed 2.6 seconds. Therefore in the case of flow curve measurements with an inertialess bob system $\mu/KS$ or $D/\omega$ should not exceed 32.5. It is relatively easy to review the validity of previous flow curve measurements by considering the ratio $D/\omega$. In most cases the viscometer constants were so adjusted that for a top angular velocity of about 42 reciprocal seconds (400 r.p.m. of the cup) the bob would rotate for a maximum of 360 degrees. Under these conditions $D/\omega=8.6$, which corresponds to a value of 0.69 second for $t_F$. This certainly means that flow curve measurements have been made with a substantial margin of safety as far as final adjustment time of the bob was concerned. This, of course, is only correct under the assumption that the oscillatory system was overdamped. Therefore, prior to a flow curve measurement it has to be determined whether the inertia of the oscillatory system is of the correct order (Eq. (2)) to assure overdamped conditions. For low viscosity materials the inertia of the bob and its associated parts were found of the correct order for the overdamped condition only after removing the deflection indicating dial.

### PRACTICAL CONSIDERATIONS FOR NON-NEWTONIAN MATERIALS

Though Eq. (1) as mentioned before is only correct for Newtonian materials, its interpretation can be used to determine whether the decrease in deflection obtained with a thixotropic material is derived from the thixotropic breakdown of that substance or from an undesirable overshooting effect of an underdamped system. The following reasoning might be applied. If the thixotropic material has a sufficiently high viscosity at the top rate of shear at which the shearing stress is measured to overdamp the oscillatory system then it must be assumed from the nature of the thixotropic substance that it would have even a higher damping effect at lower rates of shear since its viscosity must be higher. However, because of the relatively high viscosity at the low rates of shear the time $t_F$ which is required to get the bob within 1 percent of the final position could conceivably be longer than the time interval which is used in performing the step by step flow curve measurement, even if $D/\omega$ is adjusted to be less than 32.5. When that happens the measured shearing stress for a given shearing force is lower than it should be. To illustrate the condition which might exist when measuring a flow curve of a thixotropic material the assumption can be made that the ratio $D/\omega$ for the top rate of shear is again equal to 8.6 which means that the time $t_F$ equals 0.69 second. If a further assumption is made that the viscosity for the first measuring step of the flow curve is four or five times higher than for the top rate of shear, then at this point 2.76 and 3.45 seconds, respectively, would have been required to make the bob come up to within 1 percent of its final deflection. From Eq. (15) it can be calculated that if only 2.6 seconds are allowed to perform a measurement the bob was about 1.4 and 3.2 percent, respectively, off its final value at the instant of recording. In such a case the hysteresis loop of the flow curve would be recorded smaller than the material warrants, especially at the low rate of shear.

It is unfortunate that two contradictory conditions have to be met, when running flow and time-deflection curves of thixotropic materials. Because of the thixotropic nature of the material the measurement should be made speedily to obtain the flow characteristics of the materials before breakdown takes place. On the other hand due to the high viscosity in the beginning of motion the damping is so great that breakdown will start before the final deflection is reached. From Eq. (16) it can be seen that the swing in time ($t_F$) is proportional to the viscosity ($\mu$) and inversely proportional to the torsion constant ($K$). Therefore, to determine the viscosity values of thixotropic materials at low rates of shear with a higher degree of accuracy it is suggested to use a stiffer bob spring than would be otherwise employed. Though the bob deflection when using a stiffer spring will be smaller, the indicated shearing stress allowing the same limited time for taking the measurement will be greater, since the final deflection will be approached faster with the stiffer spring. Unfortunately one is restricted in increasing the spring constant since finally the motion of the bob becomes too small for accurate reacings even when a recorder is used.

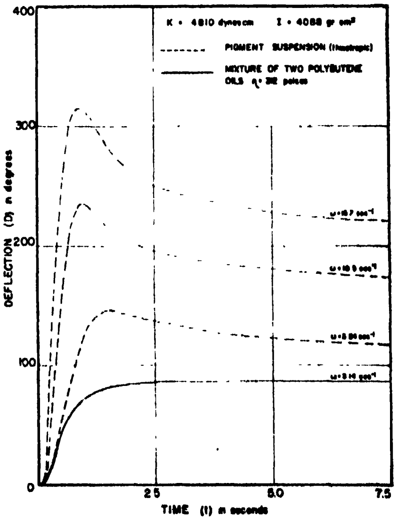


FIG. 6 Aperiodic oscillation curves of a thixotropic pigment suspension compared to that of an oil in the Newtonian region.

For the study of the initial conditions of thixotropic materials a special recorder might have to be designed which could provide some means of amplification of the purposely low deflection readings obtained with a too stiff spring. Such a recorder, for example, might make use of a long light beam which is thrown on a photographic recording paper. Another possibility might be to use a bob arrangement in which the bob is essentially motionless and the shearing force is measured with a strain gage. A high speed electrical recorder should be used to draw the flow curve from the indication of the strain gage.

### COMPARISON BETWEEN NEWTONIAN AND NON-NEWTONIAN BOB DEFLECTION-TIME CURVES

Since bob deflection-time curves can be calculated for true liquids, a deviation from the expected curve will show up any material which is non-Newtonian. Pryce Jones[5] used a method which was based on this consideration. In accordance with this procedure the bob was in a deflected position before the measurement was started and the bob deflection-time curve was recorded while the bob returned to its zero position. Recently Clark and Hodsman[6] employed a method similar to the one of Pryce-Jones for measuring viscosities.

In distinction to Pryce-Jones and Clark and Hodsman the author is presently engaged in studying at different rates of shear bob deflection-time curves which result from changes of the cup speed for Newtonian and non-Newtonian materials.

In Fig. 6 overdamped bob deflection-time curves are shown at three rates of shear for a thixotropic pigment suspension. In each case the curves were measured starting from a zero rate of shear. The breakdown in thixotropic structure is apparent from the decrease in bob deflection with time. As can be calculated by substituting the maximum deflection in Eq. (16), thixotropic breakdown occurred long before the bob would have reached its final deflection. For comparison one overdamped oil curve was added, which was taken at a rate of shear below its limiting rate of shear and which has a Newtonian viscosity of 312 poises. Attempts are being made to find some means of correlating the bob deflection values for Newtonian and non-Newtonian materials at any instant and rate of shear. No definite data are as yet available. But there are indications that this approach might make it possible to evaluate the consistency of thixotropic materials at very low rates of shear.

### ACKNOWLEDGMENT

The author is grateful to the Interchemical Corporation for permission to publish this paper and expresses thanks to Evelyn Berezin, Marjorie McElligott and Vera Osman for assistance in preparing it.

---

[5] J. Pryce-Jones, J. Oil and Col. Chem. Assn. **17**, 171 (September, 1934).

[6] A. A. Clark and H. J. Hodsman, J. Soc. Chem. Ind. **56** (February, 1937).

# Stability and High Frequency

Aurel Wintner
*The Johns Hopkins University, Baltimore 18, Maryland*

(Received April 21, 1947)

In a linear oscillator which is not conservative, let the frequency vary so as to tend to infinity when time increases indefinitely. It seems to be evident that the oscillations must then be of stable type. Actually, this is not true. The paradox results from the circumstance that the relevant criterion of stability depends on the monotone decrease, rather than on the indefinite decrease, of the instantaneous period.

## I

If a continuous $\omega=\omega(t)$, given for large positive $t$, varies so as to satisfy

$$\omega(t)\rightarrow\infty \quad \text{as} \quad t\rightarrow\infty, \tag{1}$$

it is usually assumed to be quite evident that the corresponding non-conservative oscillator

$$x''+\omega^2x=0 \tag{2}$$

must be of stable type (in the sense that

$$|x(t)|<\text{const. as } t\rightarrow\infty, \tag{3}$$

if $x=x(t)$ is any solution). A closer reflection on the energy balance shows, however, that the stipulation of (1) in (2) as a sufficient criterion for (3) is just a superstition. In fact, it turns out that the variability of $\omega(t)$ can be subject to (1) and be still so much in phase with the loss and gain in energy, that the local amplitudes of a solution of (2) fail to remain under any preassigned bound.

What *is* always true is that (2) must be of stable type if

$$d\omega(t)\geqq 0 \quad \text{when} \quad t_0<t<\infty. \tag{4}$$

But (1) does not imply that (4) is satisfied (nor is, of course, the converse inference true). In fact, if $\omega(t)$ is asymptotically proportional to $t$, say $\omega(t)=t+\theta(t)$, where $\theta(t)$ is of a lower order than $t$ as $t\rightarrow\infty$, then (1) is satisfied but (4) need not hold. For, even if the remainder term, $\omega(t)-t$, is of so small an order as to satisfy an inequality $|\theta(t)|<C$ (for some $C=$const.), possible fluctuations of $\theta(t)$ can lead to a situation violating (4). The construction of a $\omega(t)$, by means of which the actual insufficiency of the alleged stability criterion will now be proved, turns out to be precisely of this fluctuating type.

## II

The proof will consist in calculating a pair of functions, $\omega=\omega(t)$ and $r=r(t)$, which satisfy (1) and

$$r(t)\rightarrow\infty \quad \text{as} \quad t\rightarrow\infty, \tag{5}$$

respectively, and are such that

$$x(t)=r(t)\cos(t^2) \tag{6}$$

becomes a solution of (1). Clearly, (5) and (6) will imply the failure of (3); they will imply even more. In fact, (5) and (6) show that the local maxima (that is, the successive amplitudes) of the elongation $|x(t)|$, instead of being just unbounded, *tend* to $\infty$ as $t\rightarrow\infty$.

Substitution of (6) into (2) gives

$$\tfrac{1}{2}\omega^2=\{(2tr'+r)\tan(t^2)+2t^2r-\tfrac{1}{2}r''\}/r. \tag{7}$$

Hence, it is sufficient to show that it is possible to choose some function $r=r(t)$ in such a way that, on the one hand, (5) is fulfilled and, on the other hand, (7) defines, in terms of this $r=r(t)$, a *continuous* function $\omega=\omega(t)$ satisfying (1).

It is understood that $r(t)$ should have continuous derivatives. But this of itself does not take care of the proviso just italicized, since the $\tan(t^2)$, occurring in the definition of $\omega(t)$, has a sequence of infinities. These infinities can, however, be eliminated from (7), by choosing $r=r(t)$ so as to satisfy, for instance, the relation

$$(2tr'+r)r=t\cos^2(t^2). \tag{8}$$

In fact, (7) is reduced by (8) to

$$\tfrac{1}{2}\omega^2=\{t\sin(t^2)\cos(t^2)\}/r^2+2t^2-\tfrac{1}{2}r''/r; \tag{9}$$

so that the infinities of the tangent are compensated by the choice (8) of $r=r(t)$.

## III

Since the product on the left of (8) is identical with the derivative of $tr^2$, the differential equation (8) for $r=r(t)$ is satisfied by

$$r=\left\{\int_0^t s\cos^2(s^2)ds\right\}^{\frac{1}{2}}/t^{\frac{1}{2}}. \tag{10}$$

Clearly, the function (10) is asymptotically proportional to $\{t^2\}^{\frac{1}{2}}/t^{\frac{1}{2}}=t^{\frac{1}{2}}$, that is, there exists a numerical constant, say $c$, satisfying

$$r\sim ct^{\frac{1}{2}} \quad \text{as} \quad t\rightarrow\infty, \quad \text{where} \quad c>0. \tag{11}$$

But (11) implies (5). Consequently, the proof will be complete if it is verified that, when $r$ is given by (10), the (positive) function $\omega$ which is then defined by (9) is such as to satisfy (1).

According to (11), the first of the three terms on the right of (9) is asymptotically proportional to $\sin(2t^2)$ and remains, therefore, in absolute value less than a constant, as $t\rightarrow\infty$. Since the second term on the right of (9) is $t^2$, it follows that the sum of the three terms, which is $\frac{1}{2}\omega^2$, will be asymptotically equal to $t^2$ if the third term is of a lower order than $t^2$. Finally, the third term is the function $r''/r$ which, in view of (11), is of the order of $r''/t^{\frac{1}{2}}$.

Consequently, (1) will be assured if it is ascertained that $r''/t^{\frac{1}{2}}$ is of a lower order than $t^2$. This (and more) will be proved by showing that

$$r''=O(t^{\frac{1}{2}}). \tag{12}$$

In (12), use is made of the familiar notation, $f=O(t^\alpha)$, for any function $f=f(t)$ which is of the order of $t^\alpha$, *at most*. In other words, $f(t)=O(t^\alpha)$ means that the quotient $f(t)/t^\alpha$ stays between finite bounds, as $t\rightarrow\infty$. In particular, $f(t)=O(1)$, where $\alpha=0$ means that $|f|<$const. as $t\rightarrow\infty$. In the following proof of (12), the obvious rules

$$O(t^\alpha)O(t^\beta)=O(t^{\alpha+\beta}), \quad O(1)+O(1)=O(1) \tag{13}$$

will repeatedly be applied.

According to (8),

$$2tr'+r=O(t)/r=O(t^{\frac{1}{2}})$$

by (11). Since this means that $r'=O(t^{\frac{1}{2}})/t-rO(t^{-1})$, it follows that

$$r'=O(t^{-\frac{1}{2}}), \tag{14}$$

again by (11). On the other hand, differentiation of (8) gives

$$(2tr''+3r')r+(2tr'+r)r'=O(t^2).$$

Hence, from (14),

$$tr''r+O(t^{-\frac{1}{2}})r+tO(t^{-\frac{1}{2}})^2+rO(t^{-\frac{1}{2}})=O(t^2).$$

In view of (11), this can be simplified to

$$tr''r+O(1)=O(t^2).$$

Consequently, $rr''=O(t)$, and so (12) follows from (11).

## IV

This proves that condition (1), without (4), does not insure the stability of (2). On the other hand, condition (4), without (1), does insure stability (if $\omega=\omega(t)$ is positive), even though (4) is not a *necessary* condition. That it is ***sufficient*** for stability, is known to follow by an argument familiar from the theory of Bessel's functions. It proceeds as follows:

Whether (4) is satisfied or not, it is seen (by differentiation) that

$$d(x^2+x'^2/\omega^2)=x'^2d(\omega^{-2}) \tag{15}$$

is an identity in $t$ along any solution $x=x(t)$ of (2), where $\omega=\omega(t)$. If (4) is assumed, then, since $\omega(t)$ is supposed to be positive, the increment on the right of (15) cannot be positive. Hence, the same is true of the increment on the left of (15). This means that the function

$$x^2(t)+x'^2(t)/\omega^2(t) \tag{16}$$

does not increase when $t$ increases. In particular, if $t_1<t_2<\cdots$ is a sequence of $t$-values at which $x'(t)$ vanishes, then, since (16) reduces to $x^2(t)$ when $t=t_1,\ t_2,\ \cdots$,

$$|x(t_1)|\geqq|x(t_2)|\geqq\cdots.$$

In other words, the successive amplitudes $|x(t_1)|$, $|x(t_2)|$, $\cdots$, representing the relative maxima of $|x(t)|$, do not increase. This proves more than what is claimed by (3), that is, by mere stability.

# Mechanism of Fracture of Glass and Similar Brittle Solids*

Nelson W. Taylor
*Research Department, Minnesota Mining and Manufacturing Company, St. Paul, Minnesota*

(Received May 15, 1947)

A theory is proposed which connects the stress, $f$, required to break a brittle material in simple tension, with its duration of application, $t$. The slow process preceding fracture is shown to be the orientation of the atomic network contained in an elementary prism of length $r=\lambda_0E/f$, where $E$ is Young's modulus and $\lambda_0$ is the critical elongation required for fracture. The rate-controlling factor is the activation energy, $E\alpha/f$, for the orientation or rearrangement of the atomic network under the stress, $f$. Moisture on glass, and moisture plus oxygen on certain metals; are important catalytic or fatigue-promoting factors because they reduce the unit activation energy, $\alpha$. The theory leads to the equations

$$t=(1/k_0)e^{E\alpha/fkT}$$

and

$$\log t=-\log k_0+(E\alpha/2.3kT)/f,$$

where $t$ is the time for fracture (duration of the stress), $k$ is the Boltzmann constant, $T$ the absolute temperature, and $\alpha$ and $k_0$ are experimentally determined constants. The logarithmic expression has the same form as the Glathart-Preston [J. App. Phys. 17, 189 (1946)] empirical relation $\log t=-a/m+1/fm$, which, in the case of glass, appears to be valid over a time factor of $10^7$.

The theory shows why a solid object does not have a single characteristic breaking strength, and how it adjusts its fracture mechanism to whatever stress is applied. Quantitative tests of the theory are made, using fracture data on various glasses and on one glass at various temperatures. Applicability of the theory to certain aspects of fatigue of metals under stress-corrosion conditions, as well as to failure by fracture of the more rigid organic plastics, is indicated.

An equation, $E\lambda_0=2\gamma$, is proposed which connects Young's modulus and the critical fracture distance with the surface tension, $\gamma$, of the solid. Some examples are given.

## INTRODUCTION

In a discussion of the strength of "brittle" materials, it is desirable to define such terms as brittle, ductile, elastic, etc. We speak of glass as a brittle material "par excellence," and yet glass when hot is quite ductile or plastic. Under some conditions, for example under impact, tar is a brittle material, whereas if a more gentle pressure is applied over a period of time the tar will undergo some deformation, part of which is permanent and part slowly recoverable. Rubber, quite typically elastic at room temperature, is brittle at liquid-air temperatures. Thus, we see that time and temperature are important factors.

Webster's new International Dictionary, Second Edition, defines *brittle* as "breaking with very slight deformation;" *ductile* as "capable of being permanently drawn out or hammered thin; capable of being molded or worked;" *elastic* as "capable of recovering size and shape after deformation."

In view of the fact that a substance like glass is quite fluid at very high temperatures, ductile in the temperature range where it is worked into useful shapes, simultaneously ductile and elastic in the annealing range of temperature, brittle at ordinary temperatures and below, and that a substance like tar is brittle or ductile depending on the time element, it is to be seen that there is no sharp dividing line where a substance can pass exclusively from one property to another. It is not surprising, therefore, to find that a brittle material shows elastic properties including not only the instantaneous response but also lesser known reactions called elastic after-working, delayed elastic effect, primary creep, etc. Likewise, normally ductile materials, under impact or brief intense stresses show essentially brittle properties.

As soon as we recognize that "time is a necessary dimension of strength," that it is vital to any consideration of the mechanical properties, we look for the reason. The problem takes on some of the aspects of a chemical reaction running at a characteristic rate, and the concepts and methods of physical chemistry should be applicable.

The strength of brittle materials depends on many factors, not only on the time, but also on the chemical environment, on pre-stressing or residual stresses resulting from the preparation of the test piece, on the method of testing, and on

---

* Presented at the Annual Meeting, Society of Rheology, New York, New York, November 2, 1946, as part of the Bingham Memorial Symposium.

size and shape. This latter factor is very significant in thin fibers, wires, and foils. Surface cracks may also act as centers of dangerous local stresses. The common belief that every material has a characteristic breaking strength is very far from the truth.

Probably most materials are brittle at the time of actual rupture or fracture. The various molecular or atomic mechanisms which can adjust to the applied stress are exhausted just prior to the instant of fracture. "Primary creep," "slip," "delayed elastic adjustment" appear to be a necessary prelude, determining by their slowness how long the stress can be sustained.

The theory developed in this paper emphasizes this atomic rearrangement as the essential factor determining the time for fracture, and advocates a method of plotting which has both practical and theoretical value.

## I. THE EMPIRICAL LINEAR RELATION: RECIPROCAL STRESS *VS.* LOG TIME

J. L. Glathart and F. W. Preston[1] have presented a linear equation showing the strength of glass as a function of time under constant load. This equation is $1/f=a+m\log t$, where $f$ is the breaking stress, $t$ the time, and $a$ and $m$ are constants. Each line is thus defined by its slope $m=d(1/f)/d(\log t)$ and by the point $(t_0)$, where the line intersects the time axis. This equation is based on the experiments by T. C. Baker[2] in Preston's laboratory, covering the exceptionally large time range of $10^7$.

The above equation may be written

$$\log t=-a/m+1/fm.$$

The numerical values of the slope $m$ range from $0.72\times 10^{-5}$ to $2.93\times 10^{-5}$ in.²/lb. for disannealed Pyrex, dry, and for plate glass, dry, respectively. However, much smaller values of the slope are indicated in the case of glass which had been especially dried by baking at 350°C in high vacuum. From the data on experiment VII of the Baker and Preston paper the value of $m$ for soda-lime glass is calculated to be only $0.04\times 10^{-5}$. In fact, some data are shown in which there appears, within the limits of experimental error, to be no decline of strength in glass which had been very thoroughly vacuum-dried by baking at 0.1-micron mercury pressure for 10 minutes at 350°C. The value of $m$ in this case would be smaller than $0.04\times 10^{-5}$. This shows that moisture plays a very important role in bringing about fatigue or decline in strength of glass under load, with the passage of time. Removal of moisture also raises the absolute strength in experiments conducted at constant time. As stated by Baker and Preston, in certain tests the ratio of the vacuum strength to the wet strength is 2.5 for Pyrex, 2.0 for soda-lime glass, 1.7 for silica glass, and 1.5 for porcelain.

If we regard the empirical relation of Glathart and Preston as measuring the speed of a molecular process, we can consider that this speed is controlled by an activation energy, just as we find the speed of an ordinary chemical reaction controlled by such an energy. This activation energy is a minimum energy which a molecular complex must acquire before it is in a position to undergo reaction. Certain bonds must be broken or weakened before new bonds can form. It would appear that moisture has the effect of greatly lowering the activation energy for fracture of glass. In other words, the presence of moisture somehow provides a mechanism of molecular rearrangement which requires lower energy than if moisture is absent. This seems reasonable since $H_2O$ can convert the strong Si−O−Si bond into weak Si−OH bonds. A theoretical derivation of an equation taking account of this activation energy will now be set forth.

## II. THEORY OF THE FRACTURE PROCESS

We shall imagine that the elementary process in fracture consists in the slow elongation of the strongest bonds in an elementary or unit rectangular prism having length, $r$, cross-sectional area, $A$, and volume $V=rA$. Under tension $f$ parallel to the length, there will be, according to Hooke's law, an extension of the prism in the amount $\lambda_0$ where $\lambda_0/r=f/E$, $E$ being Young's modulus. This establishes the length $r$ in terms of a characteristic unit of length, $\lambda_0$. The prism may be regarded as made up of $E/f$ layers, each $\lambda_0$ in

---

[1] J. L. Glathart and F. W. Preston, "The fatigue modulus of Glass," J. App. Phys. 17, 189 (1946). See also F. W. Preston, "Strength of glass and duration of stressing," Nature 156, 55 (July 14, 1945).

[2] T. C. Baker and F. W. Preston, "Fatigue of glass under static loads; effect of water on the strength of glass," J. App. Phys. 17, 170 (1946).

thickness and $A$ in area. Let the extension $\lambda_0$ of the prism be just sufficient for fracture. The activation energy for fracture is then equal to $\alpha E/f$, where $\alpha$ is the activation energy required for orientation and associated stretching of each layer of thickness $\lambda_0$ and area $A$. (See Fig. 1.) The product $\alpha E/f$ represents the whole activation energy for orientation and consequent stretching of all of the network of atoms contained in the volume $V$. Since $\lambda_0/r=f/E$ and since $\lambda_0$ is a characteristic constant length, $r$ becomes less as $f$ increases, so that for constant cross-sectional area the volume $V$ diminishes as $f$ increases. It thus appears that the activation energy for orientation and stretching of the material contained in volume $V$ will diminish as $f$ increases. The rate of the fracture process is, therefore, the rate at which the material in $V$ acquires its critical activation energy $\alpha E/f$, and this can be written:

$$\text{rate of fracture}^{**}=k_0e^{-E\alpha/fkT}.$$

The corresponding time for fracture then becomes:

$$t=(1/k_0)e^{E\alpha/fkT}$$

and

$$\log t=-\log k_0+(E\alpha/2.3kT)/f.$$

This equation is of exactly the same form as that of the empirical Glathart-Preston relation: $\log t=-a/m+1/fm$. We see that $m=2.3kT/E\alpha$. It is to be emphasized that the assumption of a characteristic or critical energy which must be supplied per unit of length $\lambda_0$ (or more exactly, per unit of volume $\lambda_0 A$) leads to a theoretical equation exactly identical in form with the empirical linear relation, namely that the logarithm of time for fracture is inversely proportional to the stress.

The equation $m=2.3kT/E\alpha$ is dimensionally correct because $m$ has the dimensions of reciprocal stress, $E$ has the dimensions of stress, and $\alpha$ and $kT$ have the dimensions of energy. It is probably impossible to predict the value of $\alpha$ from fundamental considerations, since it is the experience in chemical kinetics that the activation energy can be calculated from fundamental constants only in the case of very simple gas reactions. However, $\alpha$ can be derived from the experimental data by the use of the equation given here.

The basic assumptions may be re-stated as follows: The strongest atomic bond in the system must be stretched a distance $\lambda_0$ in order to effect fracture. This elongation $\lambda_0$ is constant, regardless of the stress. The length of the chain of atoms which cooperates to produce this elongation $\lambda_0$ is $r=\lambda_0 E/f$ and, therefore, is inversely proportional to $f$. The elongation of this chain, or more properly speaking, the elongation of a prism of length $r$ and area $A$, is a slow process because of the many bond angles which must be changed and the many atoms which have to be squeezed and rearranged in order to permit elongation of the strongest bond in the amount $\lambda_0$. When this elongation has occurred, it can be imagined as being distributed in a statistical way among all the bonds whose total length is $r$, but because of extremely rapid thermal vibration, the whole $\lambda_0$ soon concentrates on some one bond and then fracture occurs. This redistribution is a very rapid process. The rate-controlling process is a relatively slow orientation of the whole network of atoms whose total length is $r$. The activation energy necessary for the orientation process has a characteristic value per unit of length, so that if for convenience $\lambda_0$ be regarded as that unit, we define $\alpha$ as the activation energy required per unit of length $\lambda_0$, with cross-sectional area $A$. The whole activation energy, then, for the elementary process of fracture, is $\alpha E/f$. One can see that when the stress is small, the time for fracture is

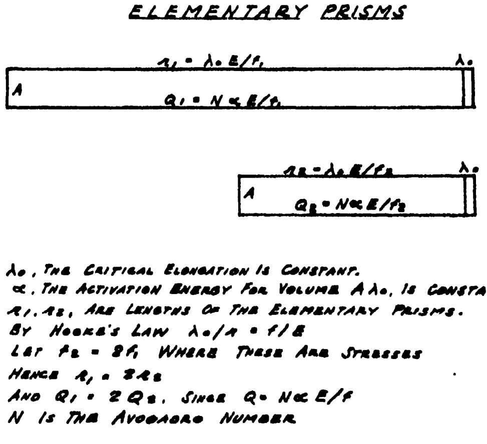

ELEMENTARY PRISMS
$r_1=\lambda_0 E/f_1$
$\lambda_0$
A
$Q_1=N\alpha E/f_1$
$r_2=\lambda_0 E/f_2$
$\lambda_0$
A
$Q_2=N\alpha E/f_2$
$\lambda_0$, THE CRITICAL ELONGATION IS CONSTANT.
$\alpha$, THE ACTIVATION ENERGY FOR VOLUME $A\lambda_0$, IS CONSTANT.
$r_1, r_2$, ARE LENGTHS OF THE ELEMENTARY PRISMS.
BY HOOKE'S LAW $\lambda_0/r=f/E$
LET $f_2=2f_1$ WHERE THESE ARE STRESSES
HENCE $r_1=2r_2$
AND $Q_1=2Q_2$, SINCE $Q=N\alpha E/f$
N IS THE AVOGADRO NUMBER

FIG. 1. Elementary prisms functioning as units in the fracture process.

---

** For details as to the connection between rate of elongation and rate of fracture, see the next section.

very long because the elementary cell or prism which has to acquire the critical activation energy $\alpha E/f$ is large. When the breaking stress is high, the elementary cell is smaller, $\alpha E/f$ is smaller, and fracture occurs sooner. The solid responds automatically to any tension which is imposed, and will break under that stress if it is applied for the right length of time. *Thus we cannot speak of a definite characteristic breaking stress of the specimen.* However, for practical purposes the times for fracture under mild stresses may be so extremely long that they can be regarded as unlimited. This is why the concept of an "endurance limit" has been found useful in the fatigue testing of metals.

### A. Phenomena Preceding Fracture

A series of researches by Taylor[3] and coworkers have shown that in the annealing range of temperature the elastic deformation of a glass fiber or rod under tension may be represented by one or more terms, each of which follows the law $dl/dt=-kl$, where $l$ is the unstretched length or the distance from equilibrium. The elongation is rapid at first but falls off with time because the fraction of the molecular network available for orientation is rapidly decreased. In this respect the phenomenon is much like radioactive decay.

The same law is found to hold at room temperature, for example, in the change of readings of a fixed point such as the ice-point of a thermometer. (See Taylor and Noyes.[4]) In one case, measured at room temperature, the "half-life" for the process was 6.75 months.

In fracture phenomena, our tests are usually of much shorter duration, and we find that the rate of elongation appears to follow the law $dl/dt = -k_0'e^{-E\alpha/fkT}$, which is to say, that for a given value of the stress, $f$, the rate $dl/dt$ is constant. Such a result would follow if the magnitude of the elongation $\lambda_0$ which is necessary for fracture is quite small compared with the original length, so that $l$ on the right side of the expression $dl/dt = -kl$ (or the apparent value of $r$ in the expression $dr/dt=-kr$) changed very little. Now the ratio $f/E$ in the glass fracture studies of Baker and Preston never exceeded 1/300, so this represents the maximum ratio of $\lambda_0/r$ for these experiments. This ratio is also small in most fatigue tests on metals, since $E$ is about $10^7$ p.s.i., and $f$ rarely exceeds $5\times10^4$, yielding a maximum $f/E$ of 1/200. Hence, we appear to be justified in converting the differential rate, $dl/dt$, to the integral form $\Delta l/\Delta t=\lambda_0/t$, where $\lambda_0$ is the characteristic critical elongation required for fracture, and $t$ is the duration of the stress.

Hence if the rate of elongation of the unit cell or prism of length, $r$, under stress, $f$, is $dr/dt = -frk_0''e^{-E\alpha/fkT}$, then $\lambda_0/t=-frk_0''e^{-E\alpha/fkT}$. But the apparent value of $r$ does not change appreciably, and $fr=E\lambda_0$, so that

$$\lambda_0/t=-E\lambda_0k_0''e^{-E\alpha/fkT},$$

$$1/t=-Ek_0''e^{-E\alpha/fkT},$$

$$t=-\frac{e^{E\alpha/fkT}}{k_0},\quad\text{where}\quad Ek_0''=k_0,$$

and

$$\log t=-\log k_0+(E\alpha/2.3kT)/f,$$

(in converting to logarithms we discard the negative sign in the preceding equation since this came originally from the equation $dr/dt=-kr$, and has no effect on the *numerical* value of $t$ or $\log t$). Negative values of $t$, the duration of stress, are meaningless.

The proper value of $E$ to use in the calculation is the mean value of the modulus over the whole range of strain from $dr/r=0$ to $dr/r=\lambda_0/r$. Since $\lambda_0/r$ is 1/200 or less we are justified in using the ordinarily measured value of $E$ which is obtained for lower values of $dr/r$ (or of $dl/l$).

If fracture tests are conducted at relatively high ratios of $f/E$, e.g., as high as 10 percent, the ratio $\lambda_0/r$ would be as high as 10 percent when fracture occurs. Some very strong short glass fibers, having a diameter $5\times10^{-4}$ inch, which were studied by F. O. Anderegg[5] showed 6 percent elongation at fracture. The percentage elongation was less for longer fibers, following a systematic trend. Anderegg estimated that a fiber of zero length would elongate 14.5 percent before fracturing. For these relatively large elongations

---

[3] N. W. Taylor, "Elastic after-effects and dielectric absorption in glass," J. App. Phys. 12, 753 (1941).

[4] N. W. Taylor and B. Noyes, "Aging thermometers," J. Am. Ceram. Soc. 27, 57 (1944).

[5] F. O. Anderegg, "Strength of glass fiber," Ind. Eng. Chem. 31, 290 (1939).

we must express the rate of elongation

$$dr/dt = -frk_0''e^{-E\alpha/fkT},$$

in the form

$$d\ \ln r/dt = -fk_0''e^{-E\alpha/fkT},$$

$$[-\ln r/(r+\lambda_0)]/t = fk_0''e^{-E\alpha/fkT}.$$

Again we discard the negative sign, so that

$$t = \frac{\ln r/(r+\lambda_0)}{fk_0''}\cdot e^{E\alpha/fkT}.$$

Actually these very strong fibers broke immediately, as soon as the full stress was reached, and it is practically impossible to determine the time effect. The reason for the extremely rapid fracture is that the ratio $E/f$ in Anderegg's case is only about 10 instead of about 300–1000 of Preston *et al*, so that the activation energy, $E\alpha/f$ is reduced by the factor 10/300 or 10/1000. This change occurs in the exponent so that it (profoundly) reduces the time for fracture, $t$, to a time which is inaccessible by our ordinary methods of measurement. The following excerpts are from Anderegg: "It seems unlikely that variation in the rate of loading is of any importance in these results. The stress-strain curves were all perfectly straight, as far as the methods would indicate." This latter statement indicates that the ordinary modulus $E$ is applicable right up to fracture.

We are therefore justified in writing our theoretical equation connecting breaking stress, $f$, with its duration, $t$ as follows:

$$\log t = -\log k_0 + (E\alpha/2.3kT)/f.$$

Some very careful experiments have been made which have failed to reveal evidence of creep. For example, F. W. Preston[6] quotes some experiments by J. T. Littleton on strong rods of borosilicate glass where very high stresses were used and where delicate optical tests failed to show creep. Preston also mentions experiments by a French worker, Grenet, in 1899, which had been carried to the point of fracture, but no displacement of optical interference fringes were observed during the last two or three minutes before fracture took place.

One possible interpretation of the negative results of Littleton and of Grenet would be that there is an internal molecular rearrangement taking place in the glass at constant total strain. In other words, the effect of the load is to produce deformation of a certain magnitude, and then with the passage of time molecular rearrangement occurs, resulting in fracture. There is direct evidence for such molecular rearrangement in glass at room temperature. Some of the early work by F. Kohlrausch[7] in 1863 demonstrated this. Kohlrausch twisted a glass filament through a given angle and measured the torque necessary to hold the filament at that constant angle of twist over a period of 24 hours. The torque gradually fell approximately proportionally to the logarithm of the time. There is much subsequent work demonstrating this relaxation phenomenon under constant deformation, and it has been shown that this is an elastic phenomenon rather than a viscous flow or permanent set, because the phenomenon is reversible. It therefore appears that we are justified in our belief that slow internal molecular rearrangements do, in fact, precede the actual fracture. Not very much detail can be given as to the actual molecular process, but it seems likely that the weaker parts of the network take up the initial strain and that this strain is transferred slowly to the stronger bonds, so as to effect a more equitable distribution of the strain energy between all the kinds of bonds in the system. However, when these stronger bonds become stretched sufficiently, fracture results.

Similar ideas regarding the transfer of strain energy by molecular rearrangement have been expressed by S. A. Shorter.[8] The following is quoted from Shorter: "If a given force be applied rapidly and maintained for a lengthy period of time, we get a rapid extension followed by a slow one. The process of extension proceeds for a very lengthy period—days or even weeks. The explanation of this is not as might be supposed that the elastic elements are showing a plastic yield, but

---

[6] F. W. Preston, "The time factor in the testing of glassware," J. Am. Ceram. Soc. **18**, 220 (1935).

[7] F. Kohlrausch, "Ueber die Elastische Nachwirkung bei der Torsion," Pogg. Ann. Physik [4] **29**, 337 (1863).

[8] S. A. Shorter, "Investigation of the nature of elasticity of fibers," J. Textile Inst. **15**, T207 (1924); "The physical properties of textile fibers in relation to technical processes and to general colloid theory," Trans. Faraday Soc. **20**, 228 (1924).

that the fiber contains elastic elements with very different degrees of damping so that on the first application of an external force, the more lightly damped elements extend, and as time goes on, the extension of the more highly damped elements begins to show itself, We get a similar effect on removing the external force, and it is undoubtedly the extreme slowness of the recovery of the more highly damped elements that has given rise to the erroneous ideas of plasticity and permanent strains. Similar considerations apply to the case where a fiber is held stretched to a definite length. We get an apparent elastic relaxation which, however, is very different from the effect contemplated in Maxwell's theory of viscosity."

A familiar example of decrease in tension in a fiber held at constant length is the decline in pitch of a violin string.

"It is not the disappearance of a state of strain owing to molecular readjustment, it is merely a transference of a state of strain from lightly damped to highly damped elements."

### B. Role of the Chemical Environment

One can only guess at the details of the atomic rearrangements which occur during the instantaneous and the delayed elastic elongation. The instantaneous part probably corresponds to elongation of the network as a whole, that is, each atom or ion is displaced or deformed slightly without disrupting the framework. This establishes a higher potential energy for the system, and *sets the stage* for the slow rearrangement which follows. This slow step is of a much more profound nature than the first since it calls for the expenditure of a substantial activation energy, which may be as high as 100,000 to 200,000 calories.[9] On the other hand, it may be quite small if "catalytic" agents such as moisture or other chemically active gases are present. In soda-lime glass which has been vacuum-baked to remove moisture, the atomic rearrangement probably involves first breaking of Na$-$O bonds, these being the weakest bonds in the structure. The stress is then transferred to stronger bonds such as Ca$-$O, and ultimately to Si$-$O which is the strongest of all. Finally, when the Si$-$O bonds are stretched to a sufficient extent, fracture occurs. In silicate crystals cleavage usually takes place in such a way as to avoid breaking the strong Si$-$O bonds, but in quartz and in glasses this cannot be avoided. All of this rearrangement (prior to fracture) could take place without any elongation beyond the original instantaneous stretch, since the stretch due to bond breaking might be compensated by a contraction due to the deformed atoms returning in part to their normal shape. We have a good deal of experimental evidence of such compensation in the behavior of electric condensers made of glass,[8] and also in the elastic response of glass, rubber, etc. as a result of a complex history.

Corrosive or chemically active gases or liquids definitely accelerate the rearrangement of the atoms of the solid and shorten the time for fracture. They do this by providing a mechanism of low activation energy, as shown by the steep $E/f$ *vs.* $\log t$ curves, the slope being $2.3kT/\alpha$. There are various possible mechanisms for this acceleration. One is that strong primary bonds such as Si$-$O$-$Ca or Si$-$O$-$Si are disrupted by reaction (e.g., with $H_2O$) and that the heat of this reaction supplies part of the requirements of the activation energy for the delayed elastic process. Likewise in metals, the heat of oxidation would supply energy to aid in the dissociation of the bonds which join the metal atoms.

The linear relation, $\log t = -a/m + 1/fm$ implies constancy of the slope $m$ over the whole duration of the test. This means that the "catalyst" has a constant influence over the whole period, and that the activation energy for molecular rearrangement is constant over this period. One would expect this to happen in glass if the moisture content remained constant. However, the situation might arise in long time tests where the sample gradually dried out. In this case, the activation energy would gradually rise and the later points of the curve would fall below the projected straight line which represented the results of the early part of the experiment. Likewise, an increase in moisture content, due to a higher humidity in the atmosphere, might cause an upward trend in the $1/f$ values. This sensitivity to the environment must be a general phenomenon in the fatigue testing of materials of

---

[9] N. W. Taylor, E. McNamara, and J. Sherman, "A study of the elastico-viscous properties of a soda-lime-silica glass at temperatures near the transformation point," J. Soc. Glass. Tech. 21, 61 (1937).

TABLE I. Percentage elongation at rupture (or fracture).

| $f$ (p.s.i.) | $E$ (p.s.i.) | calculated $(f/E)100$ | observed $(dl/l)100$ |
|---|---|---|---|
| $0.204\times 10^6$ | $2.8\times 10^6$ | 7.3 | $6.1\pm 1.7$ |
| 0.219 | 4.9 | 4.5 | $4.5\pm 0.8$ |
| 0.182 | 5.2 | 3.5 | $4.1\pm 0.2$ |
| 0.160 | 6.3 | 2.5 | $2.6\pm 0.5$ |
| 0.112 | 4.7 | 2.4 | $2.3\pm 0.4$ |
| 0.107 | 6.3 | 1.7 | $1.7\pm 0.2$ |

all kinds. The test piece should be seasoned before the fatigue tests are begun in order that it may come into "equilibrium" with the environment.

Solids under tension are at a higher potential energy and are therefore more susceptible to chemical attack than when stress-free, so that even the mildest agents may shorten the time for fracture. For this reason, investigations pointing toward more chemically-resistant materials, toward stable protective coatings, and toward reduction of tension by pre-stressing, are very much worth while. In high temperature service, corrosion may be very rapid and the danger of fracture very greatly enhanced. The problem is a critical one for the development of gas turbines, jet engines, etc.

In view of the intimate relation which appears to exist between the slow atomic rearrangement and the time required for fracture, it would be of great interest to determine quantitatively the rate constants for the delayed elastic effects in solids as affected by different chemical environments, comparing these with corresponding data from fatigue tests under stress-corrosion conditions. These studies could take the form of measurement of the decrease in Young's Modulus as a function of the time, for example by measuring the force needed to maintain a given displacement. Another procedure would be to measure the residual displacement (and its rate of relaxation) after holding a sample under a given tension for various lengths of time. Recent developments in strain gauges should make these latter measurements quite easy.

## C. Percentage Elongation at Fracture as a Test of the Theory

The present theory assumes that $\lambda_0$ is constant. The percentage elongation at fracture, $(\lambda_0/r)100$, should therefore not be constant for different stresses, but should be equal to the corresponding percentage of the breaking stress in terms of the modulus of elasticity, $(f/E)100$. Now $\lambda_0/r$ cannot be measured directly but $\lambda_0/r=dl/l$, if the tension is the same along the whole length, $l$, of the specimen. Fortunately, all the necessary data are available from the work of Anderegg,[5] who broke his very strong glass fibers by straight pulling. Table I shows Anderegg's directly observed percentage elongations and also those calculated as the ratio of his observed $f$'s and $E$'s. The data are from Table II of Anderegg's paper.

It is to be noted that the agreement is quite good. The lack of constancy in the percentage elongation forces us to discard the old idea that fracture occurs when a certain percentage elongation of the specimen is reached. On the other hand, the results are consistent with the concept that the elongation $\lambda_0$, necessary for fracture, is constant. Unfortunately, we are not in a position as yet to calculate $\lambda_0$ with certainty. Anderegg estimated by extrapolation that a fiber of zero length should elongate 14.5 percent before fracture. We might regard this as 14.5 percent of the normal length of the strongest bond in a glass, which is the silicon-oxygen bond, known from x-ray data to be 1.62 Ångstrom units. In this case, $\lambda_0$ would be $1.62\times 0.145=0.23$A. Hence, for example, when $E/f=r/\lambda_0=1000$, we find $r=230$A.

## D. Possibility of Viscous Flow?

The term "primary creep" is used in this paper in the sense of a reversible delayed elastic process. The question arises as to whether true viscous flow accompanies the elastic deformation preceding fracture of brittle materials, as it does, for example, in glasses in their annealing range of temperature. Anderegg's observations that the percentage elongation at fracture of his glass fibers were equal to the calculated elastic distortion $f/E$, would seem to eliminate the presence of viscous flow, or at least make it negligible. Under "weaker" stresses, where $f/E$ is 1/200 or less, it seems even less probable. Viscous flow is not the essential process in fracture, although it is barely possible that it goes on in parallel with the elastic deformation that does lead to fracture. Possibly one reason why a material is brittle is that it may have lost that capacity for viscous flow which it possessed at higher temperatures.

### E. Surface Tension of the Solid

The product $E\lambda_0$ measures the energy per unit area which is released when fracture occurs. If this all goes into new surface we may write $E\lambda_0=2\gamma$, where $\gamma$ is the surface tension, or specific surface energy. Taking $E=5.0\times 10^6$ p.s.i., which is the mean value of Anderegg's moduli in his Table II, and using $\lambda_0=0.23\times 10^{-8}$ cm, gives $\gamma=400$ ergs/cm². Since this result depends very much on the values assigned to $E$ and to $\lambda_0$, it must be taken with caution. It is, however, of the right order of magnitude. Parmelee, Lyon, and Harmon[10] give $\gamma=336-0.0671t$, where $t=$ °C, for a glass containing 10 percent $Na_2O$, 20 percent $B_2O_3$, 70 percent $SiO_2$, the measurements being made at 1150–1350°C. Extrapolation to 24°C yields $\gamma=335$ ergs/cm².

In cases where $E$ and $\gamma$ are known, $\lambda_0$ can be calculated, $\lambda_0=2\gamma/E$. Let us assume that the ratio $\gamma/E$ for ice is the same as that for water. From the velocity of sound in water, $v=1.45\times 10^5$ cm/sec., we calculate $E$ from the relation $v=(E/\rho)^{\frac{1}{2}}$ where $\rho$ is the density. Thus $E=2.10\times 10^{10}$ dynes/cm². Taking $\gamma=73$ dynes/cm we find $\lambda_0=0.7$A. This is about one-half the length of the H−O bond in water or ice, 1.275A. We shall make similar calculations for lead and zinc in order to get some idea of the critical bond elongation in these metals. The calculations are only approximate, since the pertinent data are not all available for one and the same temperature. For *lead*, $\gamma_{\text{liq}}=445$ dynes/cm, $E$ (calc. from the velocity of sound and density of the solid Pb)$=1.60\times 10^{11}$ dynes/cm². Then $\lambda_0=0.56$A. The Pb-Pb distance from x-ray data is 3.5A, so that $\lambda_0$ is 16 percent of this bond length. For *zinc*, $\gamma_{\text{liq}}=750$ dynes/cm, $E_{\text{solid}}=9.8\times 10^{11}$ dynes/cm². Therefore, $\lambda_0=0.15$A. The Zn-Zn distance is 2.65A, so that $\lambda_0$ is 6 percent of this bond length. The percentages are substantially less than that for $H_2O$, and more like that for the glass. Probably each substance has its characteristic percentage determined by its valences, type of binding, etc.

TABLE II. (Temperature 75°F, 297°A). Unit activation energies, $\alpha$.

| Material | $m$(in.²/lb.) | $E$(lb./in.²) | $\alpha$ |
|---|---|---|---|
| Disannealed Pyrex, dry | $0.72\times 10^{-5}$ | $8.86\times 10^6$ | $10.7k$ |
| Scratched Pyrex, wet | 1.74 | $8.86\times 10^6$ | $4.44k$ |
| Annealed soda-lime glass, wet | 1.51 | $(1\ \times 10^7)$ | $4.53k$ |
| Annealed lead glass, wet | 1.34 | $8.7\ \times 10^6$ | $6.10k$ |
| Fused silica, wet | 0.70 | $1.01\times 10^7$ | $9.67k$ |
| Porcelain, type *A*, dry | 0.88 | $1.28\times 10^7$ | $6.07k$ |
| Porcelain, type *B*, dry | 0.90 | $1.28\times 10^7$ | $5.94k$ |
| Porcelain, type *C*, dry | 1.35 | $1.28\times 10^7$ | $3.96k$ |
| Plate glass, dry | 2.93 | $1.00\times 10^7$ | $2.33k$ |

### F. Testing Procedure

Some tests on breaking strength or on fatigue are conducted with a steadily increasing load. Such tests are difficult to interpret because of the difficulty of integrating the effects of the load over a period of time. Some tests, for example, on organic plastics, are conducted under such a condition that the cross section of the sample decreases markedly during the test. These tests are also difficult, if not impossible, to interpret. Some tests apply a periodic or cyclic stress, the mean value of which may or may not be calculable. The simplest method of testing which will give results useful from a theoretical point of view are those conducted at constant stress.

The condition of the surface of the material, of course, plays an important role in determining its strength characteristics. Presence of scratches or notches act as stress concentrators and, therefore, make it difficult to calculate the actual stress existing prior to failure. Probably one of the main factors in causing scatter of test results is just this inability to be able to determine actually the stress which exists at the point where fracture starts. Precompression of the surface by special heat treatments or, in the case of metals, by peening, tends to raise the strength. However, if the amount of pre-stressing is not accurately known, the net tensional stress existing will not be truly known.

In view of the well-known influence on strength of the size and shape of the specimen, it is necessary that this be recognized in any series of tests. James Bailey[11] has given a statistical treatment

---

[10] C. W. Parmelee, K. C. Lyon, and C. G. Harmon, "The surface tensions of molten glasses," Univ. of Ill. Eng. Exp. Sta. Bull. Ser. No. 311 (1939).

[11] J. Bailey, "An attempt to correlate some tensile strength measurements on glass," Glass Ind. **20**, 21, 59, 95, 143 (1939).

which shows that if surface defects (flaws) are distributed randomly, the probability of obtaining a flaw in the highly stressed region of the specimen becomes a function of the stressed area. His idea explains the fact that the mean strength of a number of rods or laths is higher when tested in cross-bending by the single knife-edge method than by the double knife-edge method. In some A.S.T.M. experiments made at Corning, these mean values were 15,000 and 10,000 p.s.i., respectively. The latter method establishes a high uniform stress over a large area, whereas the single knife edge produces a high stress gradient.

Bailey saw that his theory was not sufficient to account for the very high strength of glass fibers, and he postulated that possibly there was a very good orientation of the "flaws" in these fibers, parallel to the fiber axis. Later, Powell and Preston[12] showed that equally high strengths could be obtained on large pieces of glass under certain conditions of loading. These authors concluded that the assumption of oriented flaws was of no help in explaining the high strengths. The writer makes an alternate proposal in this regard: exceptionally high strengths always appear to be associated with tests which place some part of the specimen in compression. Thus, for example, fine wires or fibers are compressed radially while being stretched axially, and the strength goes up as the diameter diminishes. Likewise, in the Hertz ball method[12] the area directly beneath the ball is under compression, surrounded by a cone in radial tension. The smaller the ball the higher is the measured breaking strength of the glass. Careful analysis of the stress distribution in these two cases should lead to a relation between strength and area which should be valid if the duration of the stress to produce fracture, i.e., the breaking time, is kept constant.

Bailey recognized that time is a factor in determining the breaking strength of glass, and in fact he conducted a series of tests "wherein strips of window glass were loaded by bending to a stress corresponding to approximately 12,000 p.s.i. in about 0.2 sec., then released and reloaded at a rate of about 1 cycle per sec. The samples were eventually broken by this repeated loading. The usual spread of results was found but the evidence was definite that the effect of repeated loading was, at least to some degree, cumulative. Fatigue effects could, no doubt, be entirely removed by annealing or heating the glass to a sufficient temperature to permit molecular rearrangements."

TABLE III. Molal activation energies, $Q$, for fracture.

| Material | Time ($t$) (sec.) | $f$ (p.s.i.) | $E/f$ | $Q$ (cal.) |
|---|---|---|---|---|
| Disannealed Pyrex, dry | 0.01 | 24,600 | 360 | 7,704 |
| | 1.00 | 17,100 | 518 | 11,085 |
| | 1000.00 | 12,800 | 692 | 14,809 |
| Scratched Pyrex, wet | 0.01 | 12,900 | 687 | 6,100 |
| | 1.00 | 9,000 | 984 | 8,738 |
| | 1000.00 | 5,750 | 1540 | 13,675 |
| Porcelain, type $B$, dry | 0.01 | 15,400 | 830 | 9,800 |
| | 1.00 | 11,900 | 1076 | 12,783 |
| | 1000.00 | 9,100 | 1407 | 16,715 |
| Vacuum-baked soda-lime glass | 10.00 | 25,330 | 395 | 135,090 |
| Soda-lime glass, wet | 10.00 | 10,500 | 952 | 8,625 |

## G. Importance of the $E/f$ *vs.* log$t$ Relation

The present theory indicates the desirability of plotting $1/f$ or $E/f$ *versus* log$t$. Objection may be raised to the use of reciprocal stress, $1/f$, on the ground that it is a difficult concept to understand. On the other hand, $E$ can usually be determined, and it is preferable to use $E/f$, which is the ratio of Young's modulus to the stress, and is a pure number. A plot of $E/f$ as ordinate *vs.* log$t$ as abscissa, should yield a straight line having the slope $2.3kT/\alpha$. A steep slope thus denotes a small value of $\alpha$, while a nearly horizontal line denotes that $\alpha$, the unit activation energy, is quite large. Increase of $T$ will increase the slope, if $\alpha$ remains constant.

When the object has been pre-stressed to a surface compression, $x$, parallel to the applied tension, the net tension is $f-x$. In this case a plot of $E/(f-x)$ *vs.* log$t$ should be linear, and a plot of $E/f$ *vs.* log$t$ should be concave downward. The magnitude of the original compression, $x$, can be determined by trial, until the plot of $E/(f-x)$ *vs.* log$t$ becomes linear. It may also be determined analytically by simultaneous solution of three equations, since the only unknowns are log$k_0$, $\alpha$, and $x$. In order to insure constancy of $\alpha$, the sample should be allowed to come to equilibrium with the chemical environment, before stress is applied. The existence of a substantial precom-

---

[12] H. E. Powell and F. W. Preston, "Microstrength of glass," J. Am. Ceram. Soc. 28, 145 (1945).

pression, $x$, may be responsible for the extremely long life of specimens when tested under low tensions, $f$, since $f-x$ may actually be quite small.

If a systematic study is made of the variation of $\alpha$ resulting from different chemical environments, e.g., moisture, oxygen, moisture plus oxygen, vacuum, etc. a pattern will undoubtedly develop which will lead to a more detailed understanding of the chemistry of the stress-corrosion process in any particular case, and of the means of protection. Some examples will now be presented.

## III. CALCULATION OF $\alpha$ FOR SEVERAL GLASSES

Data summarized by Glathart and Preston in their Table I "Average breaking stress in pounds per square inch (and its reciprocal) for various durations of load" permit calculation of the slopes $m$. These are shown in Table II. Additional columns show $E$ (from the International Critical Tables) and $\alpha$ calculated from the equation $m=2.3kT/E\alpha$. The value of the Boltzmann constant $k=1.38\times10^{-16}$ ergs per degree, but for convenience $\alpha$ will be expressed as a certain number of $k$'s, rather than in absolute units.

The molal activation energies, $Q$ (Table III), may be calculated for each value of the breaking stress at corresponding durations of load, by the equation $Q=N\alpha E/f$, where $N=$ Avogadro's number and where $Nk=R=2$ cal./mole. Typical calculations will be made for breaking times of 0.01 sec., 1 sec., and 1000 sec. We shall also calculate the results for a vacuum-baked (350°C) soda-lime glass where $m=0.04\times10^{-5}$ in.$^2$/lb., assuming $E$ to be $1.0\times10^7$ p.s.i. In this case $\alpha=171k$. (See Fig. 2.)

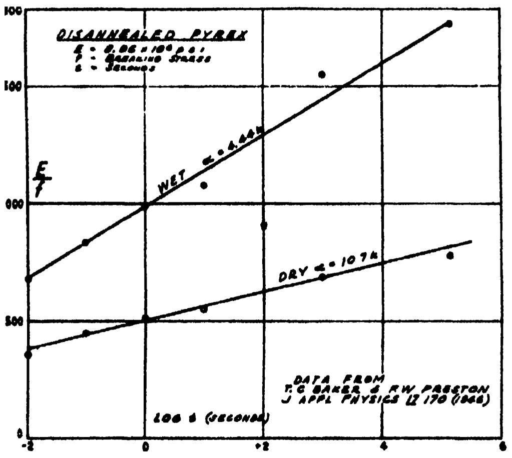


FIG. 2. $E/f$ *vs.* log$t$ for wet and dry Pyrex under static load.

It is to be noted that the molal activation energies are not large under wet conditions or "ordinary" conditions of dryness. They are of the order of 8000–16,000 cal. This can be compared with the activation energy for electrolytic or ionic conduction in soda-silica glasses (25,000 to 30,000 calories per mole.), as found by Seddon, Tippett, and Turner.[13] On the other hand, when special pains are taken to produce very dry glass, as by baking in vacuo, the molal activation energy for fracture is 135,000 cal., roughly 15 times as large as that for glass in its "ordinary" condition. The "catalytic" role of water is obvious. It provides a mechanism of low activation energy. The 135,000-calorie value is very nearly identical with 140,000 cal. for the delayed elastic process in soda-lime glass in the annealing range of temperature as found by Taylor, McNamara, and Sherman.[9] The close agreement indicates that the rate controlling process for fracture is in reality a delayed elastic phenomenon.

### A. Constancy of Young's Modulus, $E$

In the preceding calculations, it was necessary to assume that $E$ is constant for a given glass regardless of its water content. This assumption is of doubtful validity because it would seem that the replacement of strong bonds by weaker ones should lower the modulus. The writer has been unable to find the report of any scientific study of the role of water on the elastic modulus of glass. However, there is a small variation in moduli of various samples which could be attributed to varying water content. Some erratic data on vitreous silica, shown by Sosman,[14] show about twofold variation in modulus.

A reduction of $E$ would result in a corresponding increase of the calculated $\alpha$, since $\alpha=2.3kT/Em$, but, on the other hand, would

---

[13] E. Seddon, J. Tippett, and W. E. S. Turner, "Electrical conductivity of sodium metasilicate-silica glasses," J. Soc. Glass. Tech. 16, 450 (1932).

[14] R. B. Sosman, *Properties of Silica* (Chemical Catalog Company, New York, 1927), p. 450.

have no effect on the computed molal activation energies, $Q$, for fracture, since

$$Q=(E/f)N\alpha=(EN/f)(2.3kT/Em)$$
$$=(2.3RT/m)/f.$$

In fact, the molal activation energies are given directly by multiplying by $2.3RT$ the last term of the Glathart-Preston equation

$$\log t=-a/m+1/fm.$$

### B. Effect of Temperature

The effect of temperature on the strength and fatigue of soda-lime glass rods has been measured by B. Vonnegut.[15] Table IV shows values of $m$ which I have computed from that data. The last column shows $\alpha$, based on the equation $\alpha=2.3kT/Em$. $E$ is assumed to be $1\times10^7$ p.s.i. at all temperatures, for lack of better information. The value of $\alpha$ at 20°C based on Baker's work (see Table I) is $4.53k$. Vonnegut and Glathart state that they believe their "fatigue" to be too large at 24°C which leads to too small a value for $\alpha$. Similarly, $\alpha$ at 110°C is probably a little too large. There appears to be a minimum value for $\alpha$ at about 200°C. The rise in $\alpha$, that is, in activation energy, at higher temperatures is undoubtedly due to loss of moisture from the glass. At the very low temperatures the moisture is less active, possibly because of a very slow rate of diffusion into the glass at such temperature. It is to be noted that in no case, even at 520°C, does $\alpha$ rise as high as when the glass is vacuum-baked at 350°C ($\alpha=171k$). This shows how strongly moisture is retained by silicate glasses. Vonnegut and Glathart drew the conclusion that variation in the activity of adsorbed moisture on the surface of the glass is the primary cause of strength variation, and suggest that experiments should be made "over the same temperature range in high vacuum, in the absence of adsorbed films."

### C. Evaluation of Log$k_0$

The general equation

$$\log t=-\log k_0+(E\alpha/2.3kT)/f$$

permits evaluation of $\log k_0$ by determining the value of $1/f$ when $\log t=0$. In such a case $(E\alpha/2.3kT)/f=\log k_0$. The corresponding empirical equation is $1/fm=a/m$. Table V lists $1/f$, $m$, and $a/m=\log k_0$, for soda-lime glass at several temperatures, based on Vonnegut's data. No simple interpretation of the numerical values of $\log k_0$ can be given because of the variation in moisture content following the different heat treatments. The size of the rod specimens, and their surface condition as to scratches, etc., also affect $\log k_0$. A study of the breaking strength at various times and temperatures of thoroughly dry glass would be very desirable. Such measurements should be paralleled by determinations of the modulus $E$. It is also of importance to find the systematic relationship between $E$ and moisture content at various temperatures.

### IV. FRACTURE OF METALS: STRESS-CORROSION PHENOMENA

The subject of fatigue failure of metals has received a great deal of attention. References 16–27 inclusive contain data which are of interest in connection with any theory of the mechanism of fracture.[16–27] In general this work supports the

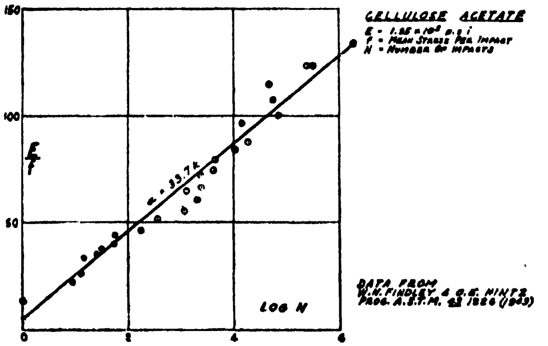


FIG. 3. $E/f$ vs. $\log N$ for cellulose acetate fractured under $N$ impacts.

---

[15] B. Vonnegut and J. L. Glathart, "The effect of temperature on the strength and fatigue of glass rods," J. App. Phys. **17**, 1082 (1946).

[16] P. W. Bridgman, "Shearing phenomena at high pressures, particularly in inorganic compounds," Proc. Am. Acad. Arts and Sci. **71**, 387 (1937).

[17] S. Glasstone, K. J. Laidler, and H. Eyring, *The Theory of Rate Processes* (McGraw-Hill Book Company, Inc., New York, 1941).

[18] E. H. Dix, Jr., "Introduction, Symposium on stress-corrosion cracking of metals," Trans. A.S.T.M. and A.I.M.E. (1946).

[19] B. P. Haigh and B. Jones, "Atmospheric action in relation to fatigue in lead," Engineering **129**, 423 (1930). Also in J. Inst. Metals, preprint No. 521.

[20] B. P. Haigh, "Chemical action in relation to fatigue in metals," Inst. Chem. Eng. (March 20, 1929).

TABLE IV. $m$ and $\alpha$ at various temperatures.

| Temperature | | $m$(in.$^2$/lb.) | $\alpha$ |
|---|---|---|---|
| −190°C | 83°A | $0.23\times10^{-5}$ | $8.31k$ |
| −80°C | 193°A | 0.33 | $13.48k$ |
| 20°C | 293°A | 2.70 | $2.50k$ |
| 110°C | 383°A | 2.89 | $3.05k$ |
| 200°C | 473°A | 4.20 | $2.59k$ |
| 300°C | 573°A | 3.10 | $4.26k$ |
| 420°C | 693°A | 1.42 | $11.2k$ |
| 520°C | 793°A | 0.94 | $19.4k$ |

idea that deformation or atomic rearrangement is a prelude to fracture and that even quite mild corrosive agents accelerate the process. However, I have not found fatigue data covering a time range which in any way approaches the $10^7$ factor which Baker was able to investigate in his glass work. The metallurgical data are particularly deficient in the short-time high stress region.

## V. FRACTURE OF ORGANIC "RESINS"

W. N. Findley and O. E. Hintz, Jr.[28] have published a paper entitled, "The relation between results of repeated blow impact tests and of fatigue tests." Their Fig. 5 is an $(S-N)$ diagram, stress *versus* log cycles, for repeated impact and fatigue-bending tests of cellulose acetate containing 26 percent plasticizer. Mr. Findley has very kindly furnished me with his actual data on number of blows at various stresses. His data can be represented by the equation: $\log N = -2.68 + 1/(0.523\times10^{-4}f)$, over the range from 10 cycles to $1.87\times10^6$ cycles. The deviation of the actual points from the straight line is no greater than the variation between comparable points at the same (or nearly the same) $\log N$. We shall take $m = 0.523\times10^{-4}$ in.$^2$/lb. Since the stress is not steady but periodic, we shall consider the mean stress over a cycle to be one-half of the recorded stress. On this basis $m = 1.05\times10^{-4}$ in.$^2$/lb. (See Fig. 3.) The unit activation energy $\alpha = 2.3kT/Em$. Findley and Hintz found $E = 1.95\times10^5$ p.s.i., and $t = 77$°F, which is 298°A. We shall use $T = 300$°A in the calculations. Thus $\alpha = 33.7k$. This is quite large compared to normal (wet) glasses where $\alpha$ ranged from $2k$ to $10k$. However, $Q$, the molal activation energy ranges from 1735 cal. (for failure in 10 impacts) to 8622 cal. (for failure in $10^6$ impacts). The corresponding values of $E/f$ are 25.7 and 128, respectively. It would be of interest to examine the relation between $\alpha$ and the kind and amount of plasticizer present.

A rough idea of $\lambda_0$ for the cellulose acetate can be obtained from the relation $\lambda_0 = 2\gamma/E$. We estimate that $\gamma$ is about 50 dynes/cm. Hence, $\lambda_0 = 0.7$A (approx.), which is about one-half of the typical C−O distance 1.42A. The typical C−C distance is about 1.54A.

This work on the fracture failure of cellulose acetate appears to demonstrate the validity of the theory when applied to a brittle organic plastic, having a modulus $E$ as low as $1.95\times10^5$ p.s.i. Since some glasses have $E$ values as high as $10^7$, a wide range of applicability is indicated.

At temperatures below zero Centrigrade, many of the common synthetic and natural organic polymers show brittle behavior. Strong threads of silk, cotton, hemp, etc., are brittle at ordinary temperatures. It would be of interest to study these materials in a quantitative way with the aid of the principles set forth in this paper.

## VI. DESIGNING STRONG PRODUCTS TO LAST

The modulus of elasticity is a fundamental constant of matter. On the other hand, it is wrong to assign one definite tensile strength to a material or even to an object, because its atoms can move and let it be broken at any stress, if the latter is applied long enough. Further, if the atmosphere or the medium has any chemical action on the material, its "strength" may decline rapidly.

---

[21] M. Fink and U. Hofmann, "Wear of metallic materials by rubbing oxidation and the question of the origin of fatigue fractures," Metallwirtschaft **13**, 623 (1934).

[22] J. N. Kenyon, "The rotating wire arc fatigue machine for testing small diameter wire," Proc. A.S.T.M. **35**, part II 156 (1935); "The fatigue properties of some cold-drawn nickel alloy wires," *ibid.* **43**, 765 (1943).

[23] F. W. Preston, "Mechanical properties of glass," J. App. Phys. **13**, 623 (1942).

[24] A. Morris, "Stress-corrosion cracking of annealed brasses," Trans. A.I.M.E. **89**, 256 (1930).

[25] W. A. Wood, "X-ray studies of grain-size in steels of different hardness values," Phil. Mag. **10**, (1930).

[26] H. J. Gough, "Crystalline structure in relation to failure of metals, especially by fatigue," Proc. A.S.T.M. **33**, 3 (1933).

[27] J. O. Almen, "The useful data to be derived from fatigue tests," Metal. Prog. 254 (1943); "Endurance of machines under a few heavy loads," *ibid.* 435 (1943); "Improving fatigue-strength of machine parts," Mech. Eng. **65**, 553 (1943).

[28] W. N. Findley and O. E. Hintz, Jr., "The relation between results of repeated blow impact tests and of fatigue tests," Proc. A.S.T.M. **43**, 1226 (1943).

The highest apparent tensile strengths are always observed where the procedure of the test places some part of the object (e.g., the surface layer) under compression. This can occur in the testing of wires, fibers, and foils, and also of more massive objects loaded on a minute area by the Hertz ball[12] (or other) method. For practical purposes the greatest progress toward highest usable strengths per unit of weight will come by paying attention to the shape factor—the use of wires, fibers, thin sheets, laminations, etc. These shapes possess a large specific surface area, which renders them more sensitive to corrosive agents. Hence, the pertinent chemistry must be well understood, and proper steps taken to make the material inert to its future environment, if this is known. The manufacturing procedures should be directed toward introducing compression into the surface region, and the product so designed and fabricated that tension stresses in service are kept to a minimum. Thus chemistry, physics, and engineering unite to form a stable tripod which carries the promise of better things for the needs of man.

TABLE V. Log$k_0$ at various temperatures.

| Temp. | $1/f$(in.$^2$/lb.) | $m$(in.$^2$/lb.) | $a/m=\log k_0$ |
|---|---|---|---|
| −190°C | $7.10\times10^{-6}$ | $0.23\times10^{-6}$ | 30.9 |
| −80°C | 8.65 | 0.33 | 26.2 |
| 20°C | 12.65 | 2.70 | 4.68 |
| 110°C | 15.40 | 2.89 | 5.33 |
| 200°C | 16.0 | 4.20 | 3.81 |
| 300°C | 14.37 | 3.10 | 4.64 |
| 420°C | 11.20 | 1.42 | 7.89 |
| 520°C | 9.27 | 0.94 | 9.86 |

## VII. ACKNOWLEDGMENTS

I wish to express my thanks to several of my colleagues in the Research Department of the Minnesota Mining and Manufacturing Company, in particular to: C. A. Dahlquist, J. O. Hendricks, G. R. Pole, A. F. Schmelzle, H. M. Scholberg, H. N. Stephens, and W. W. Wetzel, for fruitful discussions.

I am indebted also to my friends, J. O. Almen, F. O. Anderegg, H. Eyring, J. L. Glathart, F. H. MacDougall, A. Nadai, E. L. Piret, E. F. Poncelet, F. W. Preston, and H. C. Urey, for suggestions and criticism and for the stimulus arising from their genuine interest in "how things break."

# Acceleration of Electrons by a Resonant Cavity*

FRANK L. HEREFORD, JR.**
*Physics Laboratory, University of Virginia, Charlottesville, Virginia*

(Received June 16, 1947)

Electrons were accelerated by means of a single resonant cavity operating at 75-cms wave-length. Energies as high as 0.75 Mev were attained by single-stage acceleration, the electrons entering the cavity at virtually zero velocity. By turning the emergent beam about, and reinjecting it into the cavity in the opposite direction two stage acceleration was achieved and 1.25-Mev electrons were produced. Possibilities of a "shuttle accelerator" are discussed.

AMONG the various schemes currently being applied to the acceleration of charged particles is a linear accelerator[1,2] comprising a series of cavity resonators so arranged that the stream of particles passes successively through each cavity gaining energy from the high electric field in each resonator. In most of these methods, present plans specify a finite injection energy at the first cavity which in cases where this injection energy is as high as 1 to 2 Mev will require considerable apparatus.

This paper describes a single cavity electron accelerator similar to that employed by Bowen, Pulley, and Gooden[3] in which no injection apparatus is utilized, but rather the electrons enter the acceleration region at virtually zero velocity and by means of a single trip through the cavity are accelerated to relativistic energies of the order of 0.75 Mev. Such a scheme provides a simple and compact device for the production of intense pulsed beams of high energy electrons in the relativistic energy range.

A new multiple-stage acceleration method has also been successively worked out in which the beam emerging from the cavity is turned sharply through 180 degrees by an appropriate magnetic-field configuration and caused to traverse the acceleration region in the opposite direction a half cycle after the first traversal. This method offers possibilities of a "shuttle" accelerator whereby a single cavity might yield energies in the 10-Mev range.

## EXPERIMENTAL APPARATUS

The cavity employed was of the reentrant type as shown in Fig. 1. It was operated at a wave-length of approximately 75 cm, the source of high frequency power being a Western Electric 7C22 pulse operated twin triode. This tube is tunable over a 390–425-megacycle range and has a maximum peak power rating of 500 kilowatts. It was operated in this case on a four-micro-second pulse, the repetition rate being 60 cycles per second, yielding a duty cycle of 0.00024. The pulse was formed by means of a conventional spark-gap modulator with an artificial transmission line forming the square wave.

The details of the cavity construction are shown in Fig. 2. It was made from brass and later copper plated to increase the $Q$. Tuning was accomplished by adjusting the gap distance between the inner conductors, the sylphon bellows shown in Fig. 2 making this adjustment possible during operation. The radio frequency power was fed into the cavity from the concentric line output of the 7C22 oscillator by means of a magnetic coupling loop. Another sylphon-bellows arrangement made possible adjustment of this loop during operation thus facilitating proper matching of the load to the oscillator. The loop itself is indicated in Fig. 1. It was made by slipping a section of sylphon tubing inside a length of ordinary wire shielding.

## PUMPING SYSTEM

The cavity was evacuated by a Distillation Product MC275 oil-diffusion pump backed by a Megavac. The pumping lead to the cavity was rather large, being 4″ in diameter and was covered at the cavity wall by a perforated $\frac{1}{32}''$ copper plate. This plate was made to fit the

---

* This work was supported in part by a Navy Bureau of Ordnance, Contract NOrd-7873.

** National Research Predoctoral Fellow. Now at Bartol Research Foundation, Swarthmore, Pa.

[1] J. C. Slater, Phys. Rev. 70, 799A (1946).

[2] L. W. Alvarez, Phys. Rev. 70, 799A (1946).

[3] Bowen, Pulley, Gooden, Nature 157, 840 (1946).

inner wall contour of the cavity and the perforations were drilled along lines parallel to the cavity axis in order to minimize impedance to the wall currents.

By means of this system it was possible to evacuate the cavity to a pressure of $3\times10^{-6}$ mm Hg before the R.F. power was turned on. If the system had been exposed to air for an appreciable period before evacuation the pressure rose to about $4\times10^{-5}$ mm Hg in operation at resonance due to glow discharge. After several minutes of operation, however, the system cleaned up and the pressure dropped back to approximately $5\times10^{-6}$ mm Hg, where optimum acceleration conditions existed. The large 4″ diameter pumping lead to the cavity was definitely necessary in order that this condition be achieved.

### SINGLE-STAGE ACCELERATION

The source of electrons for single-stage acceleration was a 2″ length of tungsten wire wound in a helix and mounted just inside the cathode inner conductor as shown in Fig. 3(a). The filament was supported by a brass tube which slipped inside the cathode inner conductor and was adjusted by means of a sylphon bellows. Each side of the filament was coupled to the cathode inner conductor by circular mica condensers. This was necessary in order that the filament be electrically a short distance from the end of the cathode.

The essentials of the system insofar as electron optics is concerned are shown in this figure also. The gap distance of approximately 1½″ was of such length that an electron emerging at 0.75 Mev would traverse the gap in about one tenth of the 400-mc oscillator period or about $3\times10^{-10}$ second. No baffle was used over the end of the cathode inner conductor. The anode inner conductor was covered by a 0.030″ copper baffle with a ¼″ hole in the center. After traversing the acceleration region, *A*, the beam was collimated to a diameter of 0.060″, and the energy was measured by magnetic deflection, the radius of curvature being computed from the displacement on the willemite screen, *C*. The magnetic-field strength was measured by a flip coil and ballistic galvanometer and a gauss meter.

With the 7C22 operating at maximum plate voltage and the cavity properly tuned, electrons were accelerated to energies as high as 0.75 Mev. In general the extent of the spectrum was as follows:

| | |
|---|---|
| Lower limit | 0.30±10 percent Mev |
| Maximum intensity | 0.65±10 percent Mev |
| Upper limit | 0.75±10 percent Mev |

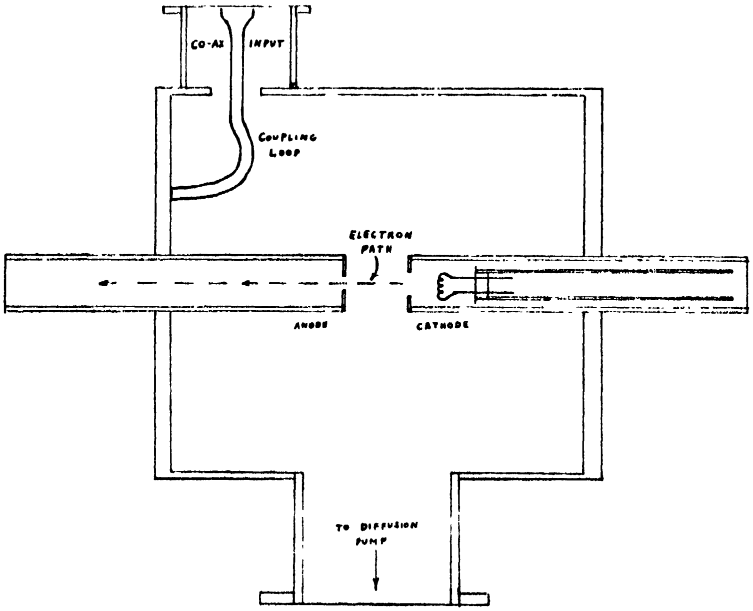


FIG. 1. A schematic drawing of the cavity as applied to single-stage acceleration.

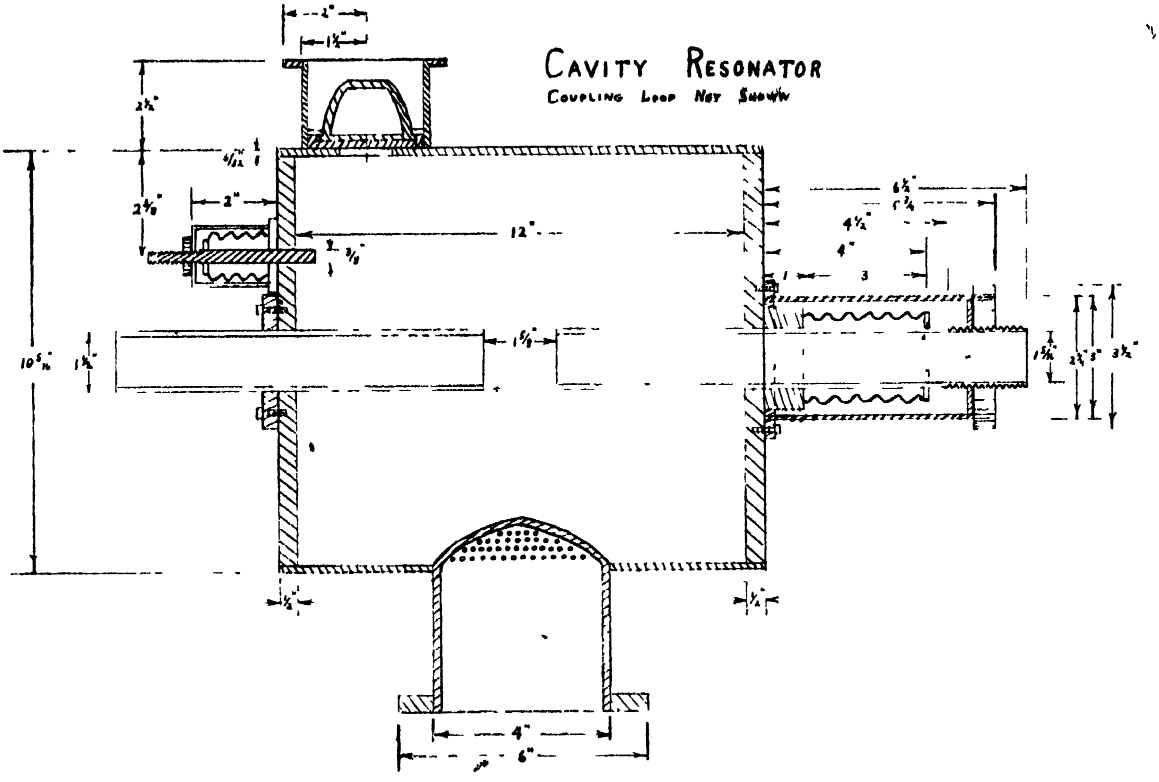


FIG. 2. Details of the cavity construction.

On several occasions an extremely narrow spectrum width was observed of the order of 50 to 100 Kev. Bowen, Pulley and Gooden[3] have reported a similar "energy bunching" effect. There seems no immediately obvious explanation of this phenomenon. A solution of the non-relativistic equations of motion of an electron in a spatially uniform electric field with a sinusoidal time variation predicts such an effect. In particular, if the energy imparted to the electron is plotted against the entrance phase angle, the curve is found to be approximately flat over a fairly large variation in phase angle. This characteristic of the curve, however, did not seem sufficiently pronounced to account for the extreme effect.

Various experiments were carried out to ascertain whether the effect was due to the focussing of only a given electron energy on the first 0.060″ aperture in the collimator, but in general changes in the electron-optics system did not effect the beam energy.

## BEAM CURRENT

By means of the scheme shown in Fig. 3(b) the current in the electron beam was measured. The beam was collimated by two $\frac{1}{4}''$ diameter apertures and electrons were collected in a Faraday cage mounted just behind the second aperture. This cage was made 7″ in length to reduce losses due to secondary electron emission from the walls.

It was possible to maintain a steady electron current to the cage of 17 microamperes. In view of the duty cycle of 0.00024 this indicated a mean beam current during the pulse of 70 milliamperes. Since the average energy was approximately 0.65 Mev, this rather high current represents a peak power loss in the $\frac{1}{4}''$ beam of 45 kilowatts. In addition to this beam there is a return beam in the opposite direction caused by the liberation of secondary electrons at the anode, some of which are emitted from the anode surface when the field is in such a direction to accelerate them. With the cavity set up for double stage acceleration as described below this return beam of secondaries was actually observed. If we assume the current in this beam to be approximately equal to that in the primary beam, we see that approximately 90-kilowatts peak power is absorbed in the electron beam alone.

The power dissipated in the cavity walls was determined by measuring the temperature rise in the walls with the R.F. power on and the cavity tuned and comparing it to the temperature rise with a heater of known dissipation in the cavity. This dissipation was found to be 220±20 kilowatts. Thus the total peak power absorbed by the cavity and the electron beam was approximately 300 kilowatts.

## DOUBLE-STAGE ACCELERATION

Using the same cavity employed in the single-stage system, double-stage acceleration was achieved by turning the beam sharply through 180 degrees as it emerged from the cavity and re-injecting it in the opposite direction. In order to carry out this scheme it was necessary to redesign the filament assembly to allow the return beam of electrons to pass through the cathode inner conductor. Furthermore, the cavity geometry was altered as indicated in Fig. 4. It was, of course, necessary that the total distance that an electron travel in emerging from the cavity, being turned about and re-injected be approximately that distance travelled in one half cycle by a 0.65-Mev electron, that being the mean energy of the beam after a single acceleration.

The double-stage system is shown in Fig. 4. The beam was not collimated after the first traversal of the gap. The two magnets used to reverse the beam direction had $\frac{3}{4}'' \times 1''$ pole faces spaced at $\frac{3}{8}''$. Magnet *A* was an electromagnet of field strength about 500 gauss, while magnet *B* was a permanent Alnico magnet of field strength about 7000 gauss. The value of $H\rho$ for 0.65-Mev electrons is 3400 gauss cms; thus the magnet *B* should turn the beam in a radius of $\frac{1}{2}$ cm.

The energy of the return beam of electrons was measured by an improved magnetic-deflection method. A wide spectrum was observed with limits as follows:

| | |
|---|---|
| Lower limit | 0.30±10 percent Mev |
| Upper limit | 1.25±10 percent Mev |

This width, however, was not due entirely to the return beam of doubly accelerated electrons. When the beam reversal magnets were removed the upper limit of the spectrum dropped to about 0.7 Mev, this beam arising from secondary emission at the anode as previously stated. Thus the lower limit of the doubly accelerated beam was approximately 0.7 Mev. The current in the return beam was considerably smaller than in the primary beam, but with improved focussing it should be possible to increase it appreciably.

There seems to be no obvious reason why such a "shuttle" acceleration method could not be applied to the production of higher energy electrons, perhaps in the 10-Mev range. With an extremely high magnetic field to reverse the beam direction the electron path in this field would be a negligible portion of the total path

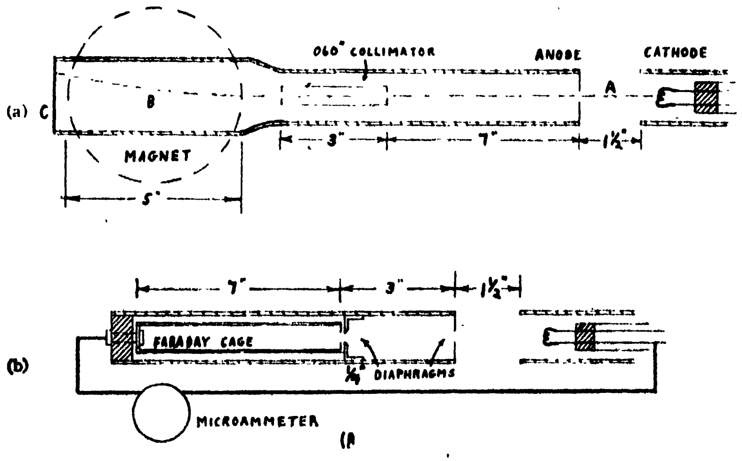


FIG. 3. (a) Schematic drawing of the single-stage system. (b) Arrangement for measurement of beam current.

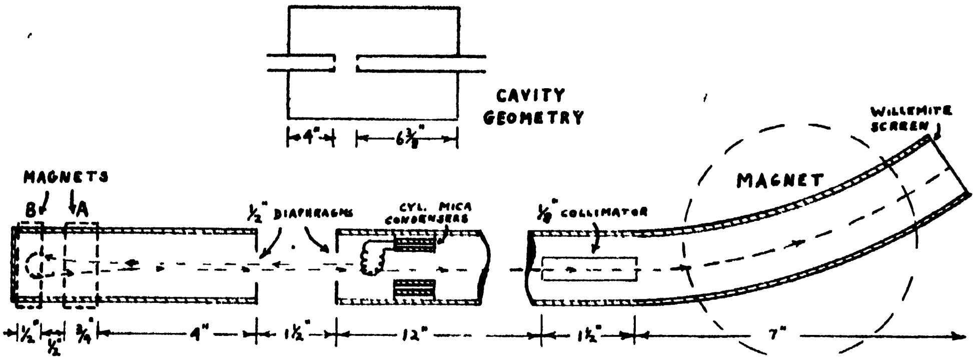


FIG. 4. The double-stage acceleration system.

per stage and consequently the increase in path due to the relativistic mass increase of the electron would not cause dephasing difficulties. Radiation losses in the reversal of the beam are of negligible value. According to an expression given by Schwinger,[4] the energy radiated in turning a 10-Mev electron about in a field of 25,000 gauss is less than 1-electron volt.

The author wishes to express his indebtedness to Professors J. W. Beams and L. B. Snoddy for the frequent and helpful advice which they gave in consultation. Their constant interest and willingness to help was sincerely appreciated. The work of Messrs. Fritz Linke and Phillip Sommer who constructed the cavity and parts also contributed much to the completion of the project.

---

[4] J. Schwinger, Phys. Rev. 70, 798A (1946).

---

# Infra-Red Spectra of Hydrocarbons

## II. Analysis of Octane Mixtures by the Use of Infra-Red Spectra Obtained at Low Temperatures*

W. H. AVERY** AND J. R. MORRISON

(Received June 17, 1947)

Infra-red spectra in the region 2 to 22 microns were obtained for the four trimethylpentanes, at temperatures of 0°C and $-195$°C. In agreement with theory, the band "widths" at $-195$°C are roughly half as great as at 0°C. The spectra were observed experimentally by condensation of a film of the hydrocarbon on the reflecting bottom surface of a cell mounted vertically, with its lower end immersed in ice or liquid nitrogen. A method of forming films of reproducible dimensions and thickness is described.

To provide a quantitative estimate of the advantage for analytical purposes of the use of spactra obtained at low temperatures, an expression is derived which gives the probability (for random distribution of the band positions) that at least one band can be found for each component of a mixture, which does not "overlap" bands of the other components. Application of the formula, with some simplifying assumptions, to a mixture of the 18 octane isomers, indicates that reduction of the band "width" from 40 $\text{cm}^{-1}$ to 20 $\text{cm}^{-1}$ (corresponding to the temperature change from 0°C to $-195$°C) increases the calculated probability of the existence of a "non-overlapped" band for at least 17 of the isomers from 0.003 to 0.999. The importance of this result in infra-red analyses is pointed out.

The use of spectra obtained at liquid helium temperatures is suggested as offering attractive possibilities for the analysis of complex mixtures.

### INTRODUCTION

THE theoretical background leading to the utilization of the low temperature technique for infra-red absorption measurements and its obvious advantages in the analytical field have been outlined previously.[1] The present paper deals with the applicability of the tech-

---

* The work reported here was completed early in 1942. Publication was delayed by the pressure of other work during the war.

** Present address is Applied Physics Laboratory, Johns Hopkins University, Silver Spring, Maryland.

[1] W. H. Avery and C. F. Ellis, J. Chem. Phys. 10, 10 (1942).

nique to the analysis of multi-component mixtures of paraffin hydrocarbons of eight carbon atoms, a complex problem of commercial as well as scientific interest since the available analytical method of precision distillation is arduous and time-consuming. Comparative absorption spectra of the trimethyl pentanes of 0°C and −195°C are presented to substantiate a theoretical discussion of the effect of band width in the problem of analyzing a complex mixture by spectroscopic methods.

## EXPERIMENTAL

The spectrometer has been described,[2] but the optical arrangement and the absorption cell mounting for low temperature measurements[1] have been revised considerably.

The present optical arrangement (Fig. 1) has the advantage of allowing the entrance of the light beam nearly normal to the cross section of the absorption cell, permitting the use of a smaller diameter cell without interception of a part of the beam. A Nernst glower, *N*, is located a few millimeters on one side of the focus of a 60° off-axis paraboloidal mirror, *P*.*** Radiation from the glower is reflected in a parallel beam from the mirror, *P*, and strikes the bottom of the cell, a plane reflecting (gold or aluminum) surface, $M_5$, at a nearly perpendicular incidence. After reflection from $M_5$ the radiation returns to the mirror, *P*, which focuses it on the plane mirror, $M_6$. $M_6$ and *N* are situated symmetrically on opposite sides of the focal point of the mirror, *P*. $M_6$, which is also at the focal point of the off-axis paraboloidal mirror $C_1$, directs the deflected beam to $C_1$. From $C_1$ the collimated beam is focused upon the entrance slit of the spectrometer by means of plane mirror $M_3$ and the condensing mirror $C_2$.

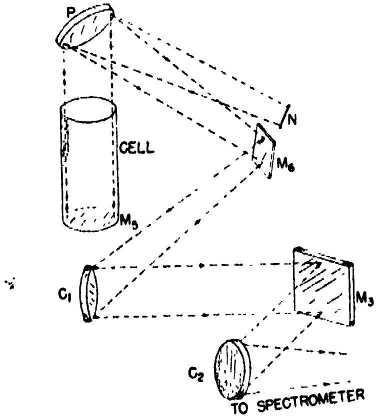


Fig. 1. Optical arrangement for low temperature spectra.

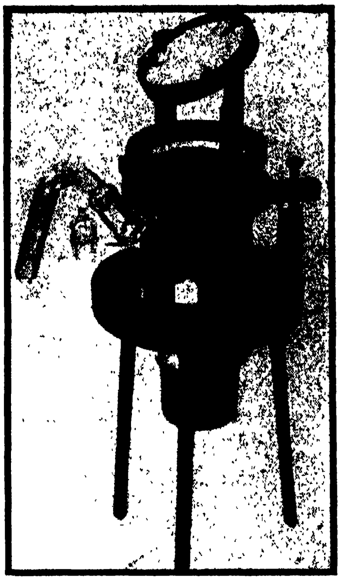


Fig. 2. Absorption cell for low temperature spectra.

The absorption cell,† 7 cm in diameter, 15 cm in depth, and fitted with a side arm for convenient sample retention during blank energy determinations, is mounted in the holder shown in Fig. 2. Both the cell holder and its supporting tripod are set into V-ways to facilitate the accurate reproduction of the cell position after filling. A similar kinematically designed arrangement permits the removal and replacement of mirror *P* without disturbing its focus. Leveling of the cell and mirror is accomplished through convenient adjustment screws.

In the original procedure a measured quantity of the sample to be examined was retained in the side arm of the absorption cell during the recording of the energy of the source in a desired spectral region. The sample was then evaporated

---

[2] W. H. Avery, J. Opt. Soc. Am. **31**, 633 (1941).

*** The mirror was prepared by the Bausch and Lomb Optical Company by cutting a section from the edge of one of their large paraboloidal searchlight mirrors.

† Obtained from the American Instrument Company.

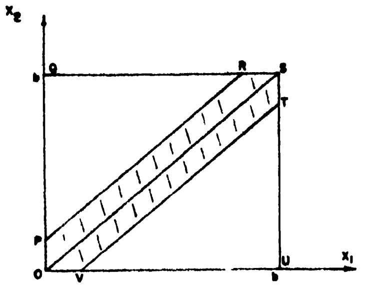


FIG. 3.

from the side arm, condensed upon the cooled cell bottom, and the same spectral region again recorded. Percentage transmissions were computed from sample to blank ratios or recorded directly by the method previously described. (A suitable factor was applied in plotting absorption curves to correct for the scattering of incident radiation from the polycrystalline film.)

The cooling of the cell bottom to −195°C was accomplished through immersion in a Dewar flask containing liquid nitrogen. By this method, however, reproducible thicknesses of deposited films were difficult to obtain since a small, variable amount of sample condensed on the cell wall adjacent to the bottom. To overcome this difficulty, a copper cylinder of a slightly smaller diameter than the cell (shown in Fig. 2) was clamped concentrically to the bottom of the cell, the cylinder rather than the cell being immersed in the liquid nitrogen. Intimate thermal contact was maintained through a thin oil film interposed between the cylinder and cell. With this arrangement the sample condensed as an accurately defined disk which did not extend to the side walls and the films were repeatedly of the same thickness to a high degree of precision. A similar technique, with the copper cylinder being immersed in an ice bath, may be used for liquid films, since films of thicknesses useful in infra-red investigations do not spread after deposition.

An absorption cell of this type presents many advantages over the conventional shim-type cell. There is no danger of contamination or evaporation of the sample; the temperature may be varied at will without condensation of water vapor or loss of sample; the sample may be recovered after examination; films of any useful thickness may be employed; the thickness may be easily varied; and, finally, the spectrum of the blank may be easily obtained.

## THE ANALYSIS OF OCTANE MIXTURES

Simplicity, accuracy, and rapidity in an infra-red analysis of a multi-component mixture depend primarily upon the location of strong absorption bands unique for each compound of interest in a spectral region relatively free from absorption by the remaining components. (An additional method has been developed[3] involving the solution, for an $n$-component system, of $n$ simultaneous equations wherein the augumentation of principal absorption by secondary absorbers caused by overlapping of bands is considered, but in this method the complexity of solution becomes increasingly great with increased values of $n$.)

Since the band widths are roughly one-half as great at −195°C as at room temperature, it is immediately obvious that the probability of locating interference-free, "characteristic" bands is increased by the utilization of the low temperature technique. A quantitative estimate of the increase is deduced in the theoretical treatment given below of the probability that an infra-red analysis can be made, taking as a criterion of possibility that at least one band of all but one of the components does not overlap a band of any other component. (It would not be necessary to find a characteristic band for each component since if all but one could be determined by the infra-red technique the last could be found by difference.) An expression is obtained applicable to any number of compounds, having any number of bands apiece and having bands of variable width from compound to compound. The position of any band is assumed to be given purely by chance.

Given $M$ compounds, designated $A$, $B$, $C$, $\cdots M$, having $n_A$, $n_B$, $n_C$, $\cdots n_M$ bands of width $w_A$, $w_B$, $w_C$, $\cdots w_M$ cm$^{-1}$ distributed at random within a spectral region $b$ cm$^{-1}$ wide, we wish to find the probability that at least one band of each $M-1$ of the compounds will occur in a region unoccupied by any bands of the other compounds.

---

[3] R. R. Brattain, R. S. Rasmussen, and A. M. Cravath, J. App. Phys. 14, 418 (1943).

Consider first the case where $M=2$, $n_A=1$, $n_B=1$. Let $x_1$ represent the position of the center of the first band $(0<x_1<b)$, $x_2$ represent the position of the band center of the second compound, $0<x_2<b$. $x_1$ and $x_2$ may be plotted as rectangular coordinates (Fig. 3); then the coordinates of every point in the square $OQSU$ will correspond to a possible disposition of the centers of the two bands. Band $A$ will overlap band $B$ whenever

$$0<x_1-x_2<\frac{w_A+w_B}{2}$$

or when

$$0<x_2-x_1<\frac{w_A+w_B}{2}.$$

$$\left(\text{Let } OP=\frac{w_A+w_B}{2}=OV=RS=ST.\right)$$

All possible values of $x_1$ and $x_2$ that satisfy the first condition lie within the shaded area of $OVTS$, while all values of $x_1$ and $x_2$ that satisfy the second condition will be found in the area $OPRS$. Let

$$\frac{w_A+w_B}{2}=w_{AB}.$$

TABLE I. 2,2,3-trimethylpentane wave-lengths and frequencies of significant bands.

| 0°C | | −195°C | |
|---|---|---|---|
| Wave-lengths in microns | Frequencies in cm⁻¹ | Wave-lengths in microns | Frequencies in cm⁻¹ |
| 19.02 | 526 | 19.12 | 523 |
| 13.95 | 717 | 13.95 | 717 |
| 12.82 | 780 | 12.89 | 776 |
| 12.07 | 829 | 12.08 | 828 |
| 11.20 | 893 | 11.23 | 890 |
| 10.77 | 929 | 10.78 | 928 |
| | | 10.37 | 964 |
| 10.30 | 971 | 10.24 | 977 |
| 9.98 | 1002 | 9.98 | 1002 |
| | | 9.81 | 1020 |
| 9.74 | 1027 | 9.72 | 1029 |
| 9,27 | 1079 | 9.26 | 1080 |
| 8.95 | 1117 | 8.93 | 1120 |
| 8.65 | 1156 | 8.65 | 1156 |
| 8.30 | 1205 | 8.30 | 1205 |
| 8.19 | 1221 | 8.19 | 1221 |
| 8.05 | 1242 | 8.02 | 1247 |
| | | 7.86 | 1274 |
| 7.65 | 1307 | 7.67 | 1304 |
| 7.30 | 1370 | 7.26 | 1377 |
| 6.77 | 1477 | 6.84 | 1462 |
| 3.83 | 2611 | 3.72 | 2688 |
| 3.50 | 2857 | 3.49 | 2865 |
| 3.26 | 3067 | 3.26 | 3067 |
| 2.34 | 4274 | 2.43 | 4115 |

The probability, $P_{AB}$, that two bands chosen at random will overlap is, then, the ratio of the area $OPRSTV$ to the area $OQSU$, i.e.,

$$P_{AB}=\frac{b^2-\dfrac{2(b-w_{AB})^2}{2}}{b^2}=1-\frac{(b-w_{AB})^2}{b^2}. \quad (1)$$

The probability that the bands do *not* overlap is

$$Q_{AB}''=1-P_{AB}=\frac{(b-w_{AB})^2}{b^2}=\left(1-\frac{w_{AB}}{b}\right)^2. \quad (2)$$

Suppose now that $A$ and $B$ have each three bands labelled $A_1$, $A_2$, $A_3$, and $B_1$, $B_2$, $B_3$' the probability that $A_1$ does not overlap $B_1$, $B_2$, or $B_3$ is

$$\left(1-\frac{w_{AB}}{b}\right)^2\left(1-\frac{w_{AB}}{b}\right)^2\left(1-\frac{w_{AB}}{b}\right)^2$$
$$=\left(1-\frac{w_{AB}}{b}\right)^6=Q_{AB}'. \quad (3)$$

TABLE II. 2,2,4-trimethylpentane wave-lengths and frequencies of significant bands.

| 0°C | | −195°C | |
|---|---|---|---|
| Wave-lengths in microns | Frequencies in cm⁻¹ | Wave-lengths in microns | Frequencies in cm⁻¹ |
| 13.45 | 743 | 13.44 | 744 |
| 12.04 | 830 | 12.03 | 831 |
| 11.59 | 863 | 11.59 | 863 |
| | | 11.09 | 902 |
| | | 10.89 | 918 |
| 10.80 | 926 | 10.77 | 929 |
| 10.19 | 981 | 10.18 | 982 |
| 9.80 | 1020 | 9.81 | 1019 |
| 9.06 | 1104 | 9.11 | 1097 |
| | | 8.97 | 1115 |
| 8.54 | 1170 | 8.55 | 1170 |
| 8.30 | 1205 | 8.28 | 1208 |
| 8.01 | 1248 | 7.99 | 1252 |
| 7.79 | 1284 | 7.80 | 1282 |
| 7.30 | 1370 | 7.35 | 1361 |
| | | 7.21 | 1387 |
| 6.78 | 1475 | 6.82 | 1466 |
| | | 4.05 | 2469 |
| 3.70 | 2703 | 3.70 | 2703 |
| 3.47 | 2882 | 3.50 | 2857 |
| 3.25 | 3077 | 3.16 | 3165 |
| 2.39 | 4184 | 3.06 | 3268 |

If $B$ has $n_B$ bands of width $w_B$,

$$Q_{AB}'=\left(1-\frac{w_{AB}}{b}\right)^{2n_B}.$$

Similar expressions are obtained for the probability that $A_2$ and $A_3$ (taken separately) do not overlap any $B$ bands.

The probability that at least one of the three $A$ bands does not overlap any of the $B$ bands will be the sum:

the probability that $A_1$ does not overlap any $B$ bands but $A_2$ and $A_3$ do
+ the probability that $A_2$ does not overlap but $A_1$ and $A_3$ do
+ the probability that $A_3$ does not overlap but $A_1$ and $A_2$ do
+ the probability that neither $A_1$ nor $A_2$ overlaps any $B$ bands but $A_3$ does
+ a similar term for $A_1$, $A_3$
+ a similar term for $A_2$, $A_3$
+, finally, the probability that neither $A_1$ nor $A_2$ nor $A_3$ overlaps any $B$ bands.

Substituting, the probability that at least one $A$ band does not overlap a $B$ band is:

$$\begin{aligned}Q_{AB}&=3Q_{AB}'(1-Q_{AB}')^2+3Q_{AB}'^2(1-Q_{AB})+Q_{AB}'^3\\&=[Q_{AB}'+(1-Q_{AB}')]^3-(1-Q_{AB}')^3\\&=1-(1-Q_{AB}')^3.\end{aligned}\qquad(4)$$

If $A$ has $n_A$ bands

$$Q_{AB}=1-(1-Q_{AB}')^{n_A}.\qquad(5)$$

If there are three compounds $A$, $B$, and $C$, the probability that at least one band of $A$ does not overlap any band in $C$ is

$$Q_{AC}=1-(1-Q_{AC}')^{n_A}.\qquad(6)$$

The probability that at least one $A$ band does not overlap any bands in either $B$ or $C$ is

$$\begin{aligned}Q_{AB}\cdot Q_{AC}&=Q_{A,B,C}\\&=[1-(1-Q_{AB}')^{n_A}][1-(1-Q_{AC}')^{n_A}].\end{aligned}\qquad(7)$$

Given $M$ compounds (where $M$ is any number), the probability that at least one $A$ band does not overlap any bands in any of the compounds is

$$Q_{A,B\cdots M}=\prod_{i=B}^{i=M}Q_{Ai}.\qquad(8)$$

Similarly, the probability that at least one $B$ band will be characteristic (i.e., not overlap) is

$$Q_{B,AC\cdots M}=\overset{(B\text{ omitted})}{\prod_{i=A}^{i=M}}Q_{Bi}.\qquad(9)$$

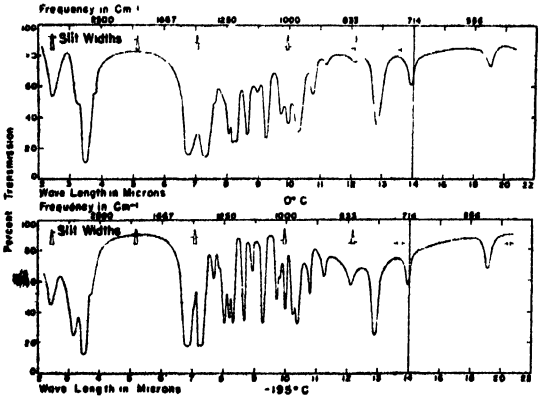


FIG. 4. Spectra of 2,2,3-trimethylpentane at 0 and −195°C.

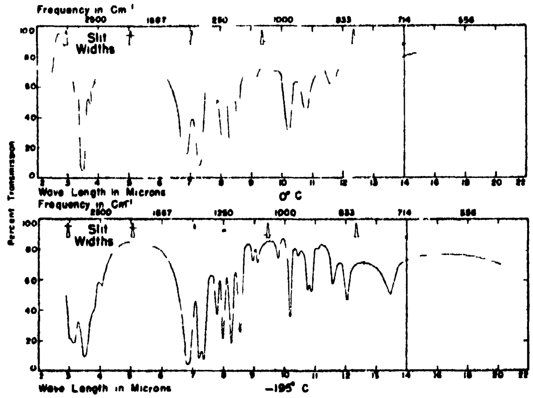


FIG. 5. Spectra of 2,2,4-trimethylpentane at 0 and −195°C.

The probability that $M-1$ compounds will have characteristic bands is††

$$P_{M-1}=\sum_{i=A}^{i=M}(1-Q_{i,A\cdots M})\prod_{j=A}^{j=M-1\ (j\neq 1)}Q_{j,A\cdots M}+\prod^{M}Q_{j,A\cdots M}.\qquad(10)$$

If the bands are all of equal width, $w$, and each

---

†† The symbol $Q_{j,A\cdot M}$ signifies that when $j$ is replaced by a given letter, that letter is to be omitted from the sequence of letters following the comma in subscript. Thus if $j=A$ we have $Q_{A,BC\cdots M}$. If $j=C$ we have $Q_{C,ABDE\cdots M}$, etc.

compound has the same number of bands, $n$, expression (10) may be simplified to

$$P_{M-1} = MQ_{A\,B\,M}^{(M-1)} - (M-1)Q_{A\,B\,M}^{M} \qquad (11)$$

$$P_{M-1} = MQ_{A,B\,M}{}^{(M-1)} - (M-1)Q_{A\,B\,.M}{}^{M} \qquad (11)$$

With these conditions,

$$Q_{AB}' = Q' = \left(1 - \frac{w}{b}\right)^{2n},$$

$$Q_{AB} = 1 - (1 - Q')^n = \left\{1 - \left[1 - \left(1 - \frac{w}{b}\right)^{2n}\right]^n\right\},$$

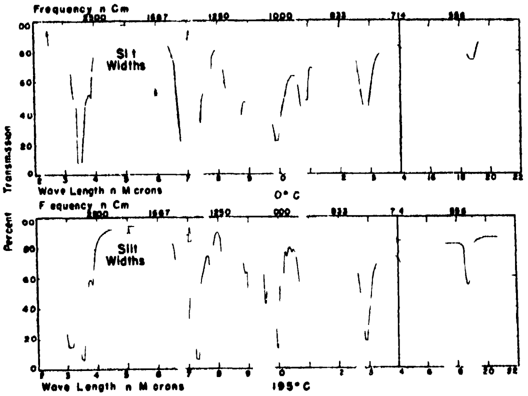


FIG. 6 Spectra of 2,3,3-trimethylpentane at 0 and −195°C

$$Q_{A\,B\,M} = \left\{1 - \left[1 - \left(1 - \frac{w}{b}\right)^{2n}\right]^n\right\}^{M-1},$$

$$P_{M-1} = M\left\{1 - \left[1 - \left(1 - \frac{w}{b}\right)^{2n}\right]^n\right\}^{(M-1)(M-1)}$$

$$-(M-1)\left\{1 - \left[1 - \left(1 - \frac{w}{b}\right)^{2n}\right]^n\right\}^{M(M-1)}$$

If $w = 0$, $P_{M-1} = M - (M-1) = 1$
If $w = b$, $P_{M-1} = 0 - 0 = 0$

If we make the assumptions, which are in approximate agreement with the facts, that all of the octane isomers have the same number of bands (15) in the region 500 to 1300 cm$^{-1}$ and that the band width at room temperature is 40 cm$^{-1}$, then: $w = 40$ cm$^{-1}$, $b = 800$ cm$^{-1}$, $n = 15$ bands, and $M = 18$ compounds

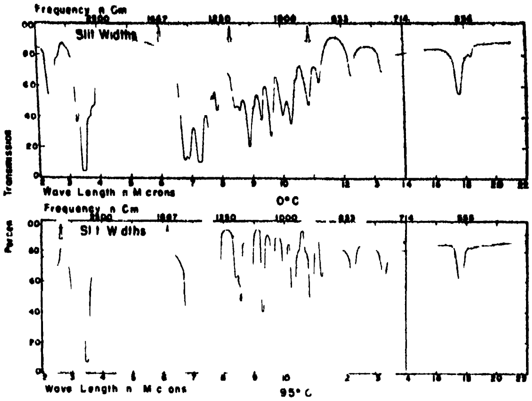


FIG. 7 Spectra of 2,3,4-trimethylpentane at 0 and −195°C.

$$Q_{A\,B\,M} = \left\{1 - \left[1 - \left(1 - \frac{40}{800}\right)^{30}\right]^{15}\right\}^{17}$$

$$= 0.628,$$

$$P_{M-1} = 18(0.628)^{17} - 17(0.628)^{18}$$

$$= 0.00268$$

When the temperature is −195°C

$$w = 40\left(\frac{78}{300}\right)^{\frac{1}{2}} = 20.4 = 20 \text{ cm}^{-1},$$

$b = 800$ cm$^{-1}$, $r = 15$ bands, $M = 18$ compounds

$$Q_{A\,B\,M} = \left\{1 - \left[1 - \left(1 - \frac{20}{800}\right)^{30}\right]^{15}\right\}^{17}$$

$$= 1 - 0.00132 = 0.99867,$$

$$P_{M-1} = 18(0.99867)^{17} - 17(0.99867)^{18}$$

$$= 0.999$$

Thus halving the band width enormously increases the probability that there will be a characteristic band for each compound.

This rather striking result has been reached through the use of the simplifying assumptions that (1) the band widths are all the same, (2) the band intensities are the same, (3) the octanes all have the same number of bands (15), and (4), that the bands are distributed at random within the interval 500–1300 cm$^{-1}$. None of the assumptions is strictly valid. However, an in-

TABLE III. 2,3,3-trimethylpentane wave-lengths and frequencies of significant bands.

| 0°C |  | −195°C |  |
|---|---|---|---|
| Wave-lengths in microns | Frequencies in cm⁻¹ | Wave-lengths in microns | Frequencies in cm⁻¹ |
| 18.82 | 531 | 18.85 | 531 |
| 14.74 | 678 | 14.90 | 671 |
|  |  | 14.36 | 696 |
| 12.87 | 777 | 12.91 | 775 |
| 12.30 | 813 | 12.08 | 828 |
| 11.18 | 894 | 11.19 | 894 |
| 10.87 | 920 | 10.88 | 919 |
| 10.74 | 931 | 10.74 | 931 |
|  |  | 10.44 | 958 |
|  |  | 10.28 | 973 |
|  |  | 10.13 | 987 |
| 9.92 | 1008 | 9.98 | 1004 |
|  |  | 9.83 | 1017 |
| 9.66 | 1035 | 9.66 | 1035 |
|  |  | 9.56 | 1046 |
| 9.19 | 1088 | 9.22 | 1085 |
| 9.02 | 1109 | 9.00 | 1111 |
|  |  | 8.89 | 1125 |
| 8.64 | 1157 | 8.61 | 1161 |
| 8.42 | 1188 | 8.40 | 1190 |
| 8.26 | 1210 | 8.23 | 1215 |
|  |  | 8.12 | 1232 |
| 7.73 | 1293 | 7.75 | 1290 |
|  |  | 7.28 | 1374 |
| 6.83 | 1464 | 6.88 | 1454 |
| 3.84 | 2604 | 3.82 | 2618 |
| 3.46 | 2890 | 3.50 | 2857 |
|  |  | 3.10 | 3226 |
| 2.36 | 4237 |  |  |

TABLE IV. 2,3,4-trimethylpentane wave-lengths and frequencies of significant bands.

| Wave-lengths in microns | Frequencies in cm⁻¹ | Wave-lengths in microns | Frequencies in cm⁻¹ |
|---|---|---|---|
| 18.40 | 544 | 18.36 | 545 |
| 17.64 | 567 | 17.59 | 569 |
| 13.28 | 753 | 13.29 | 752 |
| 12.28 | 814 | 12.26 | 816 |
| 11.22 | 891 | 11.22 | 891 |
| 10.85 | 922 | 10.89 | 918 |
|  |  | 10.70 | 935 |
| 10.45 | 957 | 10.47 | 955 |
| 10.28 | 973 | 10.29 | 972 |
| 10.01 | 999 | 10.01 | 999 |
| 9.61 | 1041 | 9.61 | 1041 |
| 9.30 | 1075 | 9.30 | 1075 |
| 9.07 | 1103 |  |  |
| 8.90 | 1124 | 8.89 | 1125 |
| 8.60 | 1163 | 8.58 | 1166 |
|  |  | 8.50 | 1176 |
| 8.43 | 1186 | 8.40 | 1190 |
| 7.86 | 1272 | 7.87 | 1271 |
| 7.72 | 1295 | 7.79 | 1284 |
| 7.59 | 1318 | 7.62 | 1312 |
| 7.26 | 1377 | 7.29 | 1372 |
| 6.90 | 1449 | 6.90 | 1449 |
| 6.78 | 1475 |  |  |
| 3.83 | 2611 | 3.83 | 2611 |
| 3.70 | 2703 | 3.70 | 2703 |
| 3.47 | 2882 | 3.49 | 2865 |
| 3.27 | 3058 |  |  |
| 3.10 | 3226 | 3.12 | 3205 |
| 2.98 | 3356 |  |  |
| 2.68 | 3731 | 2.61 | 3831 |
| 2.41 | 4150 | 2.34 | 4274 |

spection of the spectra of the trimethylpentanes presented in this paper shows that (1), (3), and (4) are approximately correct. Consideration of differences in intensities instead of (2) would seriously complicate the problem. But it would appear that the general conclusion would be unchanged, i.e., that analysis would be greatly simplified by the low temperature technique. It would appear, therefore, that use of spectra obtained at −195°C would facilitate greatly the analysis of mixtures containing all of the octanes. Analyses of mixtures of greater complexity should be possible by use of spectra obtained at the temperatures of liquid hydrogen or helium.

## SPECTRA OF TRISUBSTITUTED OCTANES

Two previously published works have contained infra-red absorption measurements over extensive regions between 2 to 20 mu for compounds of this type. In 1933, Kettering and Sleator[4] obtained the spectrum of 2,2,4-trimethylpentane in the vapor phase between 1 and 16 mu, and in 1938 Lambert and Lecomte[5] obtained the liquid spectra of all of the octanes between 7 and 15 mu. However, these measurements were made with rather low resolution and consequently are unsuitable for analytical work.

The spectra obtained in the work reported here are shown in Figs. 4 to 7 while wave-lengths and frequencies of the significant bands are listed in Tables I to IV.

The film thickness employed in each experiment was 0.071 mm, representing an absorption path of 0.142 mm (since the light traversed the film twice) and each spectrum is the result of at least four determinations for the entire region 2–22 mu. Experimental difficulties attributable to interference of absorption bands of water and carbon dioxide in the air of the laboratory prevented accurate identification of the positions and intensities of the weak hydrocarbon bands

---

[4] C. F. Kettering and W. W. Sleator, Physics **4**, 47 (1933).

[5] P. Lambert and J. Lecomte, Ann. d. Physik **10**, 503 (1938).

in the 4–6-micron region, but this region is unimportant in paraffinic surveys since the strong bands suitable for analysis occur in the region 500–1500 $cm^{-1}$.††† As a result, the curves serve only to illustrate the relative transmissions in this region.

A comparison of the positions and relative intensities of the bands at the two temperatures shows that only minor changes are produced by the change in state, with the exception of the 3-micron region. In the absence of more detailed examination of this region no explanation can be made of the pronounced bands at about 3 microns.

A comparison of the band widths in the liquid and low temperature spectra indicates close agreement between experimental and theoretical values. In nearly every case the band width at low temperature is equal (or less than) the value predicted by theory. The anticipated marked increase in resolution is also evident.

All of the pure hydrocarbons were prepared in this laboratory by the precision distillation of petroleum fractions in glass columns packed with helices. The efficiencies of the columns ranged from 60 to 100 theoretical plates. The physical properties of the samples, listed in Table V, indicate a purity of 99+ percent in every case.

TABLE V.

| Compound | *BP*, °F | $n_{20}^{D}$ | $d_4^{20}$ |
|---|---|---|---|
| 2,2,3-trimethylpentane | 229.6 | 1.40293 | 0.7160 |
| 2,2,4-trimethylpentane | 210.6 | 1.39157 | 0.6919 |
| 2,3,3-trimethylpentane | 238.6 | 1.40732 | 0.7253 |
| 2,3,4-trimethylpentane | 236.1 | 1.4045 | 0.7191 |

## CONCLUSIONS

Since a complete system of infra-red analysis cannot be developed without the spectra of all of the components for reference in locating characteristic bands, elaboration on an analytical method is reserved until the entire group of isomers has been studied. However, it is evident that such an analysis can be evolved, and through the use of diminished temperatures the method can be of a simple nature.

There exists the interesting possibility that the analyses could be further simplified by the use of liquid hydrogen or helium as cooling materials. If so, the bands would be roughly one-third to one-seventh as wide as at room temperature, assuming that the band width remains proportional to the square root of the absolute temperature.

## ACKNOWLEDGMENT

We wish to acknowledge our indebtedness to Mr. C. C. Templeton and Mr. C. F. Ellis who ably assisted in the experimental work, to Dr. H. T. Bollman, under whose direction the distillation of the samples was carried out, and to Dr. Harold Gershinowitz for his interest and advice.

---

††† In this region C–C stretching frequencies, $CH_2$, and $CH_3$ rocking and twisting, C–C–C bending frequencies, and torsional oscillations about the single bond occur. This interval is therefore most suitable for distinguishing structural differences between molecules.

# Magnetic Field Configurations Due to Air Core Coils*

JOHN P. BLEWETT**
*East Patchogue, Long Island, New York*

(Received June 17, 1947)

The field configurations around a circular loop of wire bearing current are discussed, and a tabulation is presented for the field component parallel to the axis of the loop. Formulae are derived for the current distribution which must obtain in cylindrical and ellipsoidal coils in order that the field inside be uniform. Several special cases are noted. The energy storage in the return flux is evaluated in each case.

## I. INTRODUCTION

THE designer of magnetic field coils who wishes to produce a field of a particular configuration is almost invariably hampered by the complexity of the associated mathematical analysis. If he is successful in this task, his formulae are usually so complicated that substitution in them is extremely tedious. In this paper an attempt has been made to simplify this problem, first by a partial tabulation of the fields due to a circular loop, the fundamental component of most air core coils, and, second by the calculation of several coil configurations for production of uniform fields. These basic coils can serve as components for synthesis of more complicated coil forms.

The present investigation was undertaken in connection with the design of electron accelerators having air core field coils. Since such accelerators may be operated intermittently and energy storage may be an important consideration, particular attention has been paid to the determination of energy stored in leakage fluxes.

## II. MAGNETIC FIELDS DUE TO CURRENT IN A CIRCULAR LOOP

The magnetic field configurations due to current flowing in a circular loop can be described in cylindrical coordinates using complete elliptic integrals, or in spherical polar coordinates using infinite series of spherical harmonics.[1] The latter expression converges very slowly and does not give directly the paraxial component which is usually the component of interest. We shall therefore use the former representation. The cylindrical coordinates $\rho$ and $z$ will be employed, and all quantities will be expressed in unrationalized M.K.S. units.

The radial and axial field components are as follows:

$$B_\rho=\frac{2\mu Iz}{\rho\{(a+\rho)^2+z^2\}^{\frac{1}{2}}}\left[-K+\frac{a^2+\rho^2+z^2}{(a-\rho)^2+z^2}E\right] \text{ webers per sq.m,} \quad (1)$$

$$B_z=\frac{2\mu I}{\{(a+\rho)^2+z^2\}^{\frac{1}{2}}}\left[K+\frac{a^2-\rho^2-z^2}{(a-\rho)^2+z^2}E\right] \text{ webers per sq.m,} \quad (2)$$

where

$\mu$ = permeability = $10^{-7}$ henry per meter for free space
$I$ = current in the loop in amperes
$a$ = radius of the loop in meters
$\rho$ and $z$ are expressed in meters and are measured from an origin at the center of the loop. The loop axis is the $z$ axis of the coordinate system
$K$ and $E$ are the complete elliptic integrals of the modulus

$$k^2=\frac{4a\rho}{(a+\rho)^2+z^2}. \quad (3)$$

In the regions close to the axis of the loop, or very far from the loop, $k^2$ is small, and the elliptic integrals can be expressed in the form:

$$K=\frac{\pi}{2}\left\{1+\left(\frac{1}{2}\right)^2k^2+\left(\frac{1\cdot 3}{2\cdot 4}\right)^2k^4+\left(\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\right)^2k^6+\cdots\right\}, \quad (4)$$

---

* Most of the work described in this paper was supported by the Office of Naval Research under Contract N5-ORI-178.

** Formerly Research Laboratory, General Electric Company. Now at Brookhaven National Laboratory, Brookhaven, Long Island.

[1] Compare W. R. Smythe *Static and Dynamic Electricity* (McGraw-Hill Book Company, Inc., New York, 1939), pp. 267 and 270.

TABLE I. $B_z a/\mu I$.

| $z/a$ → | 0 | 0.02 | 0.04 | 0.06 | 0.08 | 0.10 |
|---|---|---|---|---|---|---|
| $\rho/a$ ↓ 0.0 | 6.2832 | 6.2793 | 6.2682 | 6.2494 | 6.2235 | 6.1900 |
| .1 | 6.3309 | 6.3267 | 6.3153 | 6.2961 | 6.2687 | 6.2341 |
| .2 | 6.4792 | 6.4744 | 6.4612 | 6.4396 | 6.4092 | 6.3705 |
| .3 | 6.7465 | 6.7408 | 6.7254 | 6.6988 | 6.6621 | 6.6159 |
| .4 | 7.1713 | 7.1642 | 7.1429 | 7.1075 | 7.0593 | 6.9976 |
| .5 | 7.8264 | 7.8160 | 7.7847 | 7.7323 | 7.6613 | 7.5715 |
| .52 | 7.9962 | 7.9847 | 7.9501 | 7.8930 | 7.8150 | 7.7164 |
| .54 | 8.1825 | 8.1699 | 8.1316 | 8.0684 | 7.9820 | 7.8744 |
| .56 | 8.3873 | 8.3733 | 8.3303 | 8.2602 | 8.1652 | 8.0454 |
| .58 | 8.6132 | 8.5971 | 8.5492 | 8.4712 | 8.3646 | 8.2312 |
| .60 | 8.8631 | 8.8453 | 8.7913 | 8.7031 | 8.5829 | 8.4344 |
| .62 | 9.1409 | 9.1201 | 9.0588 | 8.9586 | 8.8228 | 8.6549 |
| .64 | 9.4500 | 9.4267 | 9.3557 | 9.2412 | 9.0866 | 8.8960 |
| .66 | 9.7966 | 9.7695 | 9.6884 | 9.5560 | 9.3775 | 9.1595 |
| .68 | 10.1886 | 10.1568 | 10.0609 | 9.9060 | 9.6989 | 9.4467 |
| .70 | 10.6329 | 10.5944 | 10.4808 | 10.2987 | 10.0551 | 9.7622 |
| .72 | 11.1401 | 11.0944 | 10.9583 | 10.7404 | 10.4515 | 10.1060 |
| .74 | 11.7272 | 11.6696 | 11.5051 | 11.2404 | 10.8932 | 10.4834 |
| .76 | 12.4097 | 12.3398 | 12.1334 | 11.8090 | 11.3864 | 10.8916 |
| .78 | 13.2160 | 13.1269 | 12.8661 | 12.4582 | 11.9347 | 11.3350 |
| .80 | 14.1831 | 14.0644 | 13.7293 | 13.2076 | 12.5481 | 11.8081 |
| .82 | 15.3580 | 15.2025 | 14.7516 | 14.0670 | 13.2269 | 12.3028 |
| .84 | 16.8247 | 16.6169 | 15.9888 | 15.0683 | 13.9694 | 12.8011 |
| .86 | 18.7023 | 18.3841 | 17.4982 | 16.2175 | 14.7505 | 13.2647 |
| .88 | 21.1950 | 20.7008 | 19.3663 | 17.5167 | 15.5219 | 13.6181 |
| .90 | 24.6635 | 23.8317 | 21.6833 | 18.9180 | 16.1756 | 13.7500 |
| .92 | 29.8546 | 28.2885 | 24.5048 | 20.2236 | 16.4402 | 13.4517 |
| .94 | 38.3974 | 34.9057 | 27.6377 | 20.8636 | 15.8663 | 12.4408 |
| .96 | 44.6382 | 45.1306 | 29.5790 | 19.4803 | 13.7732 | 10.4189 |
| .98 | 93.9679 | 43.8104 | 24.4478 | 13.9945 | 9.5639 | 7.2928 |
| 1.00 | ∞ | 4.9912 | 4.2969 | 3.8899 | 3.6004 | 3.3734 |
| 1.02 | −94.0877 | −44.9238 | −15.6759 | −6.1192 | −2.3052 | −.4881 |
| 1.04 | −44.8502 | −35.1530 | −20.6741 | −11.4415 | −6.3880 | −3.5412 |
| 1.06 | −28.5817 | −25.4049 | −18.8187 | −12.7635 | −8.3979 | −5.4477 |
| 1.08 | −20.6171 | −19.2334 | −15.8890 | −12.1590 | −8.9202 | −6.3894 |
| 1.10 | −15.8750 | −15.1625 | −13.3003 | −10.9663 | −8.6526 | −6.6536 |
| 1.12 | −12.7434 | −12.3320 | −11.2241 | −9.7017 | −8.0675 | −6.5359 |
| 1.14 | −10.5495 | −10.3852 | −9.5686 | −8.5542 | −7.3872 | −6.2133 |
| 1.16 | −8.9219 | −9.1774 | −8.2640 | −7.5549 | −6.6980 | −5.8115 |
| 1.18 | −7.6723 | −7.5554 | −7.2182 | −6.7013 | −6.0718 | −5.3767 |
| 1.20 | −6.6928 | −6.6012 | −6.3570 | −5.9751 | −5.4984 | −4.9640 |
| 1.22 | −5.8949 | −5.8330 | −5.6448 | −5.3561 | −4.9903 | −4.5691 |
| 1.24 | −5.2448 | −5.1971 | −5.0540 | −4.8309 | −4.5406 | −4.2059 |
| 1.26 | −4.6983 | −4.6613 | −4.5497 | −4.3749 | −4.1481 | −3.8789 |
| 1.28 | −4.2409 | −4.2123 | −4.1231 | −3.9839 | −3.7978 | −3.5792 |
| 1.30 | −3.8473 | −3.8239 | −3.7522 | −3.6398 | −3.4885 | −3.3091 |
| 1.32 | −3.5081 | −3.4892 | −3.4315 | −3.3389 | −3.2147 | −3.0630 |
| 1.34 | −3.2139 | −3.1985 | −3.1503 | −3.0746 | −2.9715 | −2.8447 |
| 1.36 | −2.9564 | −2.9413 | −2.9019 | −2.8384 | −2.7521 | −2.6455 |
| 1.38 | −2.7287 | −2.7180 | −2.6846 | −2.6290 | −2.5558 | −2.4660 |
| 1.40 | −2.5266 | −2.5156 | −2.4879 | −2.4429 | −2.3807 | −2.3034 |
| 1.42 | −2.3460 | −2.3384 | −2.3143 | −2.2742 | −2.2210 | −2.1548 |
| 1.44 | −2.1842 | −2.1782 | −2.1576 | −2.1231 | −2.0773 | −2.0188 |
| 1.46 | −2.0400 | −2.0329 | −2.0150 | −1.9853 | −1.9463 | −1.8964 |
| 1.48 | −1.9086 | −1.9023 | −1.8869 | −1.8612 | −1.8273 | −1.7825 |
| 1.5 | −1.7891 | −1.7850 | −1.7703 | −1.7490 | −1.7178 | −1.6785 |
| 1.6 | −1.3321 | −1.3290 | −1.3220 | −1.3095 | −1.2917 | −1.2713 |
| 1.7 | −1.0267 | −1.0256 | −1.0207 | −1.0139 | −1.0030 | −.9906 |
| 1.8 | −.8131 | −.8125 | −.8096 | −.8054 | −.7986 | −.7903 |
| 1.9 | −.6580 | −.6577 | −.6557 | −.6526 | −.6482 | −.6427 |
| 2.0 | −.5416 | −.5414 | −.5401 | −.5381 | −.5346 | −.5308 |

TABLE I.—*Continued.*

| $z/a$ → | 0.12 | 0.14 | 0.16 | 0.18 | 0.20 | 0.22 | 0.24 |
|---|---|---|---|---|---|---|---|
| $\rho/a$ ↓ 0.0 | 6.1497 | 6.1028 | 6.0496 | 5.9896 | 5.9243 | 5.8533 | 5.7769 |
| .1 | 6.1925 | 6.1436 | 6.0882 | 6.0271 | 5.9590 | 5.8854 | 5.8069 |
| .2 | 6.3242 | 6.2703 | 6.2087 | 6.1403 | 6.0651 | 5.9844 | 5.8977 |
| .3 | 6.5596 | 6.4942 | 6.4210 | 6.3389 | 6.2503 | 6.1547 | 6.0533 |
| .4 | 6.9242 | 6.8389 | 6.7439 | 6.6389 | 6.5255 | 6.4046 | 6.2777 |
| .5 | 7.4644 | 7.3430 | 7.2077 | 7.0606 | 6.9044 | 6.7400 | 6.5693 |
| .52 | 7.6004 | 7.4670 | 7.3206 | 7.1620 | 6.9933 | 6.8166 | 6.6344 |
| .54 | 7.7469 | 7.6013 | 7.4417 | 7.2690 | 7.0870 | 6.8968 | 6.7006 |
| .56 | 7.9055 | 7.7453 | 7.5701 | 7.3819 | 7.1841 | 6.9788 | 6.7680 |
| .58 | 8.0756 | 7.8991 | 7.7064 | 7.5011 | 7.2856 | 7.0631 | 6.8362 |
| .60 | 8.2594 | 8.0646 | 7.8512 | 7.6256 | 7.3896 | 7.1478 | 6.9034 |
| .62 | 8.4596 | 8.2412 | 8.0048 | 7.7557 | 7.4971 | 7.2337 | 6.9686 |
| .64 | 8.6748 | 8.4295 | 8.1662 | 7.8902 | 7.6055 | 7.3178 | 7.0307 |
| .66 | 8.9080 | 8.6311 | 8.3350 | 8.0281 | 7.7138 | 7.3992 | 7.0870 |
| .68 | 9.1587 | 8.8448 | 8.5117 | 8.1679 | 7.8208 | 7.4755 | 7.1357 |
| .70 | 9.4302 | 9.0686 | 8.6925 | 8.3068 | 7.9217 | 7.5424 | 7.1734 |
| .72 | 9.7189 | 9.3046 | 8.8747 | 8.4426 | 8.0140 | 7.5970 | 7.1963 |
| .74 | 10.0273 | 9.5475 | 9.0572 | 8.5682 | 8.0916 | 7.6346 | 7.2001 |
| .76 | 10.3541 | 9.7926 | 9.2300 | 8.6778 | 8.1493 | 7.6480 | 7.1787 |
| .78 | 10.6923 | 10.0349 | 9.3864 | 8.7646 | 8.1760 | 7.6298 | 7.1240 |
| .80 | 11.0666 | 10.2594 | 9.5131 | 8.8123 | 8.1626 | 7.5689 | 7.0305 |
| .82 | 11.3620 | 10.4499 | 9.5922 | 8.8052 | 8.0936 | 7.4572 | 6.8871 |
| .84 | 11.6525 | 10.5762 | 9.5984 | 8.7236 | 7.9548 | 7.2784 | 6.6850 |
| .86 | 11.8621 | 10.6012 | 9.4967 | 8.5425 | 7.7238 | 7.0195 | 6.4173 |
| .88 | 11.9247 | 10.4696 | 9.2515 | 8.2321 | 7.3826 | 6.6715 | 6.0720 |
| .90 | 11.7385 | 10.1080 | 8.8074 | 7.7622 | 6.9122 | 6.2176 | 5.6459 |
| .92 | 11.1689 | 9.4404 | 8.1253 | 7.1064 | 6.3008 | 5.6585 | 5.1361 |
| .94 | 10.0596 | 8.3880 | 7.1668 | 6.2523 | 5.5488 | 4.9978 | 4.5503 |
| .96 | 8.3201 | 6.9152 | 5.9358 | 5.2146 | 4.6709 | 4.2407 | 3.8981 |
| .98 | 5.9421 | 5.0775 | 4.4776 | 4.0349 | 3.6944 | 3.4235 | 3.2008 |
| 1.00 | 3.1949 | 3.0343 | 2.8979 | 2.7816 | 2.6715 | 2.5713 | 2.4834 |
| 1.02 | .4743 | 1.0118 | 1.3422 | 1.5358 | 1.6599 | 1.7334 | 1.7737 |
| 1.04 | −1.8235 | −.7692 | −.0727 | .3931 | .7134 | .9375 | 1.0986 |
| 1.06 | −3.4863 | −2.1580 | −1.2398 | −.5903 | −.1230 | .2237 | .4804 |
| 1.08 | −4.5162 | −3.1463 | −2.1366 | −1.3837 | −.8190 | −.3916 | −.0649 |
| 1.10 | −5.0309 | −3.6799 | −2.7532 | −1.9764 | −1.3729 | −.9010 | −.5259 |
| 1.12 | −5.1963 | −4.0669 | −3.1426 | −2.3944 | −1.7915 | −1.3007 | −.9046 |
| 1.14 | −5.1358 | −4.1771 | −3.3567 | −2.6653 | −2.0851 | −1.6037 | −1.2022 |
| 1.16 | −4.9444 | −4.1465 | −3.4364 | −2.8117 | −2.2772 | −1.8198 | −1.4297 |
| 1.18 | −4.6927 | −4.0341 | −3.4227 | −2.8735 | −2.3891 | −1.9626 | −1.5987 |
| 1.20 | −4.4116 | −3.8727 | −3.3544 | −2.8742 | −2.4384 | −2.0507 | −1.7054 |
| 1.22 | −4.1250 | −3.6734 | −3.2371 | −2.8249 | −2.4417 | −2.0925 | −1.7757 |
| 1.24 | −3.8454 | −3.4726 | −3.1035 | −2.7470 | −2.4108 | −2.0969 | −1.8100 |
| 1.26 | −3.5811 | −3.2711 | −2.9594 | −2.6525 | −2.3566 | −2.0777 | −1.8174 |
| 1.28 | −3.3340 | −3.0752 | −2.8089 | −2.5449 | −2.2863 | −2.0405 | −1.8069 |
| 1.30 | −3.1051 | −2.8854 | −2.6592 | −2.4324 | −2.2071 | −1.9880 | −1.7793 |
| 1.32 | −2.8927 | −2.7097 | −2.5167 | −2.3189 | −2.1222 | −1.9289 | −1.7413 |
| 1.34 | −2.7013 | −2.5434 | −2.3783 | −2.2069 | −2.0338 | −1.8629 | −1.6960 |
| 1.36 | −2.5242 | −2.3894 | −2.2451 | −2.0964 | −1.9451 | −1.7945 | −1.6456 |
| 1.38 | −2.3627 | −2.2466 | −2.1223 | −1.9909 | −1.8580 | −1.7233 | −1.5909 |
| 1.40 | −2.2132 | −2.1131 | −2.0052 | −1.8908 | −1.7739 | −1.6549 | −1.5351 |
| 1.42 | −2.0764 | −1.9905 | −1.8948 | −1.7959 | −1.6915 | −1.5858 | −1.4792 |
| 1.44 | −1.9525 | −1.8758 | −1.7936 | −1.7053 | −1.6132 | −1.5179 | −1.4225 |
| 1.46 | −1.8374 | −1.7703 | −1.6970 | −1.6192 | −1.5373 | −1.4534 | −1.3676 |
| 1.48 | −1.7310 | −1.6720 | −1.6088 | −1.5390 | −1.4663 | −1.3907 | −1.3131 |
| 1.5 | −1.6332 | −1.5819 | −1.5245 | −1.4638 | −1.3979 | −1.3300 | −1.2606 |
| 1.6 | −1.2489 | −1.2158 | −1.1828 | −1.1472 | −1.1084 | −1.0679 | −1.0255 |
| 1.7 | −.9751 | −.9572 | −.9371 | −.9143 | −.8905 | −.8651 | −.8373 |
| 1.8 | −.7803 | −.7687 | −.7556 | −.7412 | −.7253 | −.7086 | −.6901 |
| 1.9 | −.6363 | −.6283 | −.6196 | −.6098 | −.5990 | −.5874 | −.5750 |
| 2.0 | −.5262 | −.5208 | −.5147 | −.5080 | −.5002 | −.4923 | −.4835 |

$$E=\frac{\pi}{2}\left\{1-\left(\frac{1}{2}\right)^2k^2-\left(\frac{1\cdot 3}{2\cdot 4}\right)^2\frac{k^4}{3}-\left(\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\right)^2\frac{k^6}{5}-\cdots\right\}. \tag{5}$$

If, as is usually the case, the region of particular interest is that close to the wire, it will be found that $k^2$ approaches unity. Within a distance of almost $0.5a$ of the wire, $k^2$ will be greater than 0.9. For $k^2$ greater than 0.9, the approximate expressions:

$$K=\ln 4-0.5\ln(1-k^2)$$
$$=1.39-1.15\log_{10}(1-k^2), \tag{6}$$

$$E=1/k^2 \tag{7}$$

give values correct to within about 3 percent.

Unfortunately, no approximate expressions seem to make any very material simplification in the relations (1) and (2). If the procedures of coil designing are to be expedited, it seems that there is no alternative to a tabulation. Since the axial component $B_z$ has been of primary interest as far as air core field coils are concerned, the tabulation has been completed only for $B_z$. The

TABLE I.—*Continued.*

| $z/a$ → | 0.26 | 0.28 | 0.30 | 0.32 | 0.34 | 0.36 |
|---|---|---|---|---|---|---|
| $\rho/a$ ↓ 0.0 | 5.6963 | 5.6104 | 5.5216 | 5.4280 | 5.3325 | 5.2335 |
| .1 | 5.7234 | 5.6355 | 5.5434 | 5.4480 | 5.3500 | 5.2483 |
| .2 | 5.8066 | 5.7108 | 5.6103 | 5.5070 | 5.4003 | 5.2912 |
| .3 | 5.9470 | 5.8358 | 5.7201 | 5.6014 | 5.4810 | 5.3574 |
| .4 | 6.1452 | 6.0083 | 5.8681 | 5.7252 | 5.5805 | 5.4354 |
| .5 | 6.3940 | 6.2157 | 6.0358 | 5.8555 | 5.6761 | 5.4982 |
| .52 | 6.4477 | 6.2582 | 6.0685 | 5.8788 | 5.6904 | 5.5051 |
| .54 | 6.5012 | 6.3007 | 6.0998 | 5.8998 | 5.7020 | 5.5081 |
| .56 | 6.5555 | 6.3412 | 6.1282 | 5.9176 | 5.7101 | 5.5072 |
| .58 | 6.6077 | 6.3797 | 6.1537 | 5.9310 | 5.7133 | 5.5014 |
| .60 | 6.6580 | 6.4143 | 6.1747 | 5.9420 | 5.7113 | 5.4893 |
| .62 | 6.7047 | 6.4448 | 6.1897 | 5.9421 | 5.7022 | 5.4708 |
| .64 | 6.7471 | 6.4688 | 6.1981 | 5.9363 | 5.6846 | 5.4438 |
| .66 | 6.7813 | 6.4839 | 6.1966 | 5.9207 | 5.6577 | 5.4062 |
| .68 | 6.8066 | 6.4886 | 6.1840 | 5.8943 | 5.6188 | 5.3582 |
| .70 | 6.8188 | 6.4800 | 6.1579 | 5.8535 | 5.5667 | 5.2972 |
| .72 | 6.8152 | 6.4547 | 6.1151 | 5.7962 | 5.4991 | 5.2211 |
| .74 | 6.7913 | 6.4084 | 6.0513 | 5.7207 | 5.4133 | 5.1286 |
| .76 | 6.7415 | 6.3385 | 5.9656 | 5.6228 | 5.3080 | 5.0185 |
| .78 | 6.6618 | 6.2375 | 5.8520 | 5.5007 | 5.1802 | 4.8874 |
| .80 | 6.5435 | 6.1035 | 5.7081 | 5.3505 | 5.0274 | 4.7359 |
| .82 | 6.3804 | 5.9298 | 5.5280 | 5.1695 | 4.8484 | 4.5597 |
| .84 | 6.1668 | 5.7123 | 5.3105 | 4.9567 | 4.6423 | 4.3607 |
| .86 | 5.8967 | 5.4454 | 5.0529 | 4.7091 | 4.4068 | 4.1401 |
| .88 | 5.5623 | 5.1281 | 4.7534 | 4.4281 | 4.1431 | 3.8915 |
| .90 | 5.1649 | 4.7583 | 4.4127 | 4.1124 | 3.8522 | 3.6232 |
| .92 | 4.7087 | 4.3409 | 4.0325 | 3.7670 | 3.5361 | 3.3348 |
| .94 | 4.1830 | 3.8781 | 3.6171 | 3.3946 | 3.2003 | 3.0294 |
| .96 | 3.6156 | 3.3785 | 3 1760 | 3.0002 | 2.8478 | 2.7076 |
| .98 | 3.0136 | 2.8558 | 2.7180 | 2.5937 | 2.4859 | 2.3865 |
| 1.00 | 2.4012 | 2.3240 | 2.2513 | 2.1828 | 2.1195 | 2.0592 |
| 1.02 | 1.7924 | 1.7965 | 1.7898 | 1.7765 | 1.7590 | 1.7354 |
| 1.04 | 1.2094 | 1.2893 | 1.3460 | 1.3831 | 1.4073 | 1.4203 |
| 1.06 | .6723 | 8162 | .9271 | 1.0100 | 1.0739 | 1 1190 |
| 1.08 | .1905 | .3856 | .5429 | 6638 | .7600 | 8369 |
| 1.10 | −.2289 | .0069 | .1975 | .3504 | .4734 | .5763 |
| 1.12 | −.5841 | -.3211 | −.1066 | 0710 | .2158 | 3120 |
| 1.14 | −.8729 | −.5948 | −.3661 | - 1742 | − 0126 | .1218 |
| 1.16 | −1.1002 | −.8198 | −.5833 | −.3814 | −.2099 | −.0673 |
| 1.18 | −1.2734 | −.9988 | −.7625 | −.5583 | −.3822 | −.2312 |
| 1.20 | −1.4022 | −1.1379 | −.9052 | −.7019 | −.5268 | −.3714 |
| 1.22 | −1.4911 | −1.2388 | −1.0164 | −.8191 | −.6437 | −.4906 |
| 1.24 | −1.5476 | −1.3121 | −1.0996 | −.9086 | −.7394 | -.5898 |
| 1.26 | −1.5800 | −1.3602 | −1.1595 | −.9788 | −.8167 | −.6702 |
| 1.28 | −1.5877 | −1.3857 | −1.1994 | −1.0299 | −.8748 | −.7336 |
| 1.30 | −1.5814 | −1.3972 | −1.2248 | −1.0652 | −.9178 | −.7826 |
| 1.32 | −1.5628 | −1.3937 | −1.2355 | −1.0861 | −.9498 | −.8220 |
| 1.34 | −1.5340 | −1.3803 | −1.2342 | −1.0978 | −.9695 | −.8488 |
| 1.36 | −1.5001 | −1.3595 | −1.2263 | −1.0992 | −.9795 | −.8678 |
| 1.38 | −1.4604 | −1.3334 | −1.2104 | −1.0942 | −.9834 | −.8780 |
| 1.40 | −1.4181 | −1.3016 | −1.1902 | −1.0836 | −.9799 | −.8826 |
| 1.42 | −1.3726 | −1.2679 | −1.1663 | −1.0671 | −.9722 | −.8808 |
| 1.44 | −1.3272 | −1.2322 | −1.1384 | −1.0483 | −.9597 | −.8756 |
| 1.46 | −1.2805 | −1.1943 | −1.1094 | −1.0258 | −.9457 | −.8669 |
| 1.48 | −1.2344 | −1.1567 | −1.0783 | −1.0023 | −.9276 | −.8551 |
| 1.5 | −1.1898 | −1.1184 | −1.0473 | −.9776 | −.9082 | −.8411 |
| 1.6 | −.9809 | −.9361 | −.8902 | −.8443 | −.7984 | −.7527 |
| 1.7 | −.8093 | −.7793 | −.7494 | −.7188 | −.6873 | −.6553 |
| 1.8 | −.6715 | −.6513 | −.6309 | −.6096 | −.5879 | −.5659 |
| 1.9 | −.5614 | −.5479 | −.5338 | −.5184 | −.5030 | −.4870 |
| 2.0 | −.4740 | −.4642 | −.4538 | −.4428 | −.4318 | −.4200 |

computations involved the use of a computing machine and an interpolation table of $K$ and $E$ drawn up from Dwight's tables.[2] The results are collected in Table I in which the values of $aB_z/\mu I$ are given for $\rho/a$ between 0.0 and 2.0, and for $z/a$ between 0.0 and 0.36.

In Fig. 1, $aB_z/\mu I$ has been plotted as a function of $\rho/a$ for three values of $z/a$. The curves have a common characteristic form. For the values of $z/a$ which have been calculated, $B_z$ increases with distance from the axis to pass through a maximum at a radius less than that of the loop. It then decreases, goes through zero at a radius greater than that of the loop, and has a minimum at a point roughly as far outside the loop radius as was the maximum inside the loop radius. From the minimum, the field increases with increasing radius and tends asymptotically to zero. In Fig. 2 the radial coordinates and heights of the maxima and minima are plotted as functions of $z/a$.

If values of $B_\rho$ are desired, fairly good accuracy can be attained from a numerical integration using the table of $B_z$ and the relation

$$(\text{curl } B)_\phi = \frac{\partial B_\rho}{\partial z} - \frac{\partial B_z}{\partial \rho} = 0. \tag{8}$$

The integration proceeds from $z=0$, where $B_\rho$ is zero for all values of $\rho$. In several cases this method of determining $B_\rho$ has been found to be materially faster than substitution in Eq. (1).

## III. PRODUCTION OF A UNIFORM TRANSVERSE FIELD INSIDE AN INFINITE CYLINDER

In a cylinder of radius $R$ and infinite length, whose axis is the $x$ axis in a right-handed rectangular coordinate system, a uniform field $B_0$ in the $z$ direction will be produced by a current $I_x$ in the cylinder wall having current density[3]

$$I_x \doteq -\frac{B_0}{2\pi\mu}\frac{y}{R}. \tag{9}$$

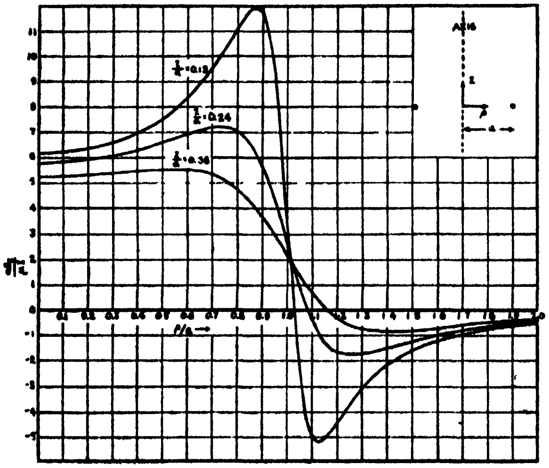


FIG. 1. Axial magnetic fields around a circular loop carrying current $I$.

---

[2] H. B. Dwight, *Mathematical Tables* (McGraw-Hill Book Company, New York, 1941), pp. 199 *et. seq.*

[3] Compare I. I. Rabi, Rev. Sci. Inst. 5, 78 (1934).

Outside the cylinder the field components will be

$$\left.\begin{aligned} B_x &= 0 \\ B_y &= 2B_0R^2\frac{yz}{(y^2+z^2)^2} \\ B_z &= B_0R^2\frac{z^2-y^2}{(y^2+z^2)^2} \end{aligned}\right\}. \tag{10}$$

The field pattern outside the cylinder is the same as that which would be produced by currents $+B_0/\delta$ and $-B_0/\delta$ parallel to the axis of the cylinder and displaced $-\delta$ and $+\delta$ from the axis in the $xy$ plane, $\delta$ being an infinitesimal.

The energy stored inside and outside the cylinder is given as usual by integrating $B^2/(8\pi\mu)$ over the appropriate volume. The energy stored per unit length in the field external to the cylinder is

$$E_0=\int_R^\infty \frac{B_0{}^2R^4\pi d(y^2+z^2)}{8\pi\mu(y^2+z^2)^2}=\frac{B_0{}^2R^2}{8\mu}. \tag{11}$$

The energy stored per unit length inside the cylinder is simply

$$E_i=\frac{B_0{}^2}{8\pi\mu}\cdot\pi R^2.$$

So in this case, $E_0=E_i$.

## IV. ELLIPSOIDAL COILS FOR PRODUCTION OF UNIFORM FIELDS

The mathematical techniques necessary for dealing with ellipsoidal coils for production of uniform magnetic fields are similar to those applying to dielectric ellipsoids in uniform electric fields.[4] We shall consider only ellipsoids of revolution in which it is desired to establish a uniform field in the direction of the axis of symmetry. In cylindrical coordinates, the equation of such an ellipsoid is

$$\frac{\rho^2}{a^2}+\frac{z^2}{b^2}=1. \tag{12}$$

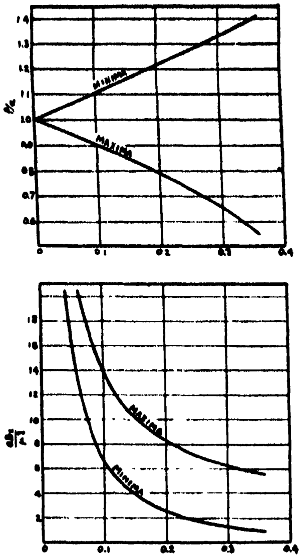


FIG. 2. Positions and heights of maxima and minima in fields of circular loop.

It is convenient to describe the external fields in terms of the parameter $s$ which describes the ellipsoids confocal with the ellipsoid of Eq. (12) through the relation

$$\frac{\rho^2}{a^2+s}+\frac{z^2}{b^2+s}=1. \tag{13}$$

The solution for $s$ of (13) is

$$s=\tfrac{1}{2}\{\rho^2+z^2-a^2-b^2\pm[(a^2+b^2-\rho^2-z^2)^2 +4(\rho^2b^2+z^2a^2-a^2b^2)]^{\frac{1}{2}}\}. \tag{14}$$

The choice of sign in (14) must be such that on the surface of the ellipsoid $s=0$ and outside of the ellipsoid $s$ is always positive. (If $s$ does not satisfy this criterion, Eq. (13) describes a family of confocal hyperboloids rather than ellipsoids.) For the most part, the criterion is satisfied by a plus sign in Eq. (14).

The external fields are derivable from a scalar potential function which, if $a$ is greater than $b$, has the form

$$V=B_0Z\left\{\frac{\dfrac{1}{(s+b^2)^{\frac{1}{2}}}-\dfrac{1}{(a^2-b^2)^{\frac{1}{2}}}\tan^{-1}\left(\dfrac{a^2-b^2}{s+b^2}\right)^{\frac{1}{2}}}{-\dfrac{b}{a^2}+\dfrac{1}{(a^2-b^2)^{\frac{1}{2}}}\tan^{-1}\dfrac{(a^2-b^2)^{\frac{1}{2}}}{b}}\right\}. \tag{15a}$$

---

[4] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, New York, 1941), Sections 3.25 to 3.27.

If $a$ is less than $b$

$$V=B_0z\left\{\frac{\dfrac{1}{(s+b^2)^{1/2}}-\dfrac{1}{(b^2-a^2)^{1/2}}\tanh^{-1}\left(\dfrac{b^2-a^2}{s+b^2}\right)^{1/2}}{\dfrac{b}{a^2}+\dfrac{1}{(b^2-a^2)^{1/2}}\tanh^{-1}\dfrac{(b^2-a^2)^{1/2}}{b}}\right\}, \tag{15b}$$

where $B_0$ is the value of the uniform field inside the ellipsoid. These functions were derived by methods similar to those used by Stratton in dealing with the dielectric ellipsoid. The fields derived from this function for the case $a$ greater than $b$, are:

$$B_\rho=-\frac{\partial V}{\partial\rho}=\frac{B_0z\rho(a^2-b^2)}{(2s-\rho^2-z^2+a^2+b^2)(s+a^2)(s+b^2)^{1/2}\left[\dfrac{b}{a^2}+\dfrac{1}{(a^2-b^2)^{1/2}}\tan^{-1}\dfrac{(a^2-b^2)^{1/2}}{b}\right]}, \tag{16}$$

$$B_z=-\frac{\partial V}{\partial z}=\frac{B_0\left\{\dfrac{z^2(a^2-b^2)}{(2s-\rho^2-z^2+a^2+b^2)(s+b^2)}-1+\left(\dfrac{s+b^2}{a^2-b^2}\right)^{1/2}\tan^{-1}\left(\dfrac{a^2-b^2}{s+b^2}\right)^{1/2}\right\}}{(s+b^2)^{1/2}\left[\dfrac{b}{a^2}+\dfrac{1}{(a^2-b^2)^{1/2}}\tan^{-1}\dfrac{(a^2-b^2)^{1/2}}{b}\right]}. \tag{17}$$

If $a$ is less than $b$, the changes necessary in (16) and (17) are obvious from a comparison of (15a) and (15b).

From the potential function of (15a) or (15b) we can deduce also the field components $B_n$ and $B_t$ normal and tangent to the surface of the ellipsoid:

$$B_n=\frac{B_0za^2}{(\rho^2b^4+z^2a^4)^{1/2}} \tag{18}$$

$$B_t=\frac{B_0\rho b^2}{(\rho^2b^4+z^2a^4)^{1/2}}\left\{\frac{\dfrac{1}{b}-\dfrac{1}{(a^2-b^2)^{1/2}}\tan^{-1}\dfrac{(a^2-b^2)^{1/2}}{b}}{\dfrac{b}{a^2}-\dfrac{1}{(a^2-b^2)^{1/2}}\tan^{-1}\dfrac{(a^2-b^2)^{1/2}}{b}}\right\}. \tag{19}$$

Inside the ellipsoid, the normal field component is the same as that outside given by (18). The tangential field component $B_t'$ at the inside surface of the ellipsoid is

$$B_t'=\frac{B_0\rho b^2}{(\rho^2b^4+z^2a^4)^{1/2}}. \tag{20}$$

From Maxwell's equations, the azimuthal current density

$$I_\phi=\frac{B_t-B_t'}{4\pi\mu}=\frac{B_0}{4\pi\mu}\frac{\rho b^2}{(\rho^2b^4+z^2a^4)^{1/2}}\left\{\frac{a^2-b^2}{b^2-\dfrac{ba^2}{(a^2-b^2)^{1/2}}\tan^{-1}\dfrac{(a^2-b^2)^{1/2}}{b}}\right\}. \tag{21}$$

This is the current per unit distance measured tangent to the surface of the ellipsoid. Between $\rho$

and $\rho+d\rho$ the current will be

$$I(\rho)d\rho = I_\phi d\rho[1+(dz/d\rho)^2]^{\frac{1}{2}} = \frac{B_0 d\rho}{4\pi\mu}\frac{\rho b}{a(a^2-\rho^2)^{\frac{1}{2}}}\left\{\frac{a^2-b^2}{b^2-\dfrac{ba^2}{(a^2-b^2)^{\frac{1}{2}}}\tan^{-1}\dfrac{(a^2-b^2)^{\frac{1}{2}}}{b}}\right\}. \tag{22}$$

Similarly, between $z$ and $z+dz$ the current will be

$$I(z)dz = \frac{B_0 dz}{4\pi\mu}\left\{\frac{a^2-b^2}{b^2-\dfrac{ba^2}{(a^2-b^2)^{\frac{1}{2}}}\tan^{-1}\dfrac{(a^2-b^2)^{\frac{1}{2}}}{b}}\right\}, \tag{23}$$

which is a constant independent of $z$. From this fact we may conclude that if the current distribution is achieved by varying the thickness of the ellipsoid wall, keeping constant current density within the wall, the wall thickness in the radial direction will be a constant. It is assumed that the wall thickness remains infinitesimal compared with the ellipsoid dimensions; for thick walls the above theory is no longer correct.

## V. ENERGY STORAGE IN FIELDS OF ELLIPSOIDAL COILS

The energy stored in the field outside the ellipsoid is evaluated most easily by an application of Green's theorem. The energy is given by

$$E_0 = \int \frac{B\cdot B}{8\pi\mu}d\tau = -\int \frac{V_s B_n}{8\pi\mu}dS, \tag{24}$$

where $V_s$ is the value of $V$ (from 15a or 15b) on the surface of the ellipsoid, and $B_n$, the normal field just outside the surface, is given by (18). The integral is taken over the surface of the ellipsoid. When the appropriate substitutions have been made in (24), the integral is easily evaluated. The result is

$$E_0 = \frac{a^2bB_0^2}{6\mu}\left\{\frac{\dfrac{1}{b}-\dfrac{1}{(a^2-b^2)^{\frac{1}{2}}}\tan^{-1}\dfrac{(a^2-b^2)^{\frac{1}{2}}}{b}}{\dfrac{b}{a^2}-\dfrac{1}{(a^2-b^2)^{\frac{1}{2}}}\tan^{-1}\dfrac{(a^2-b^2)^{\frac{1}{2}}}{b}}\right\}. \tag{25}$$

The energy inside the ellipsoid where the field is uniform is

$$E_i = \frac{4}{3}\pi a^2 b\frac{B_0^2}{8\pi\mu} = \frac{a^2bB_0^2}{6\mu}. \tag{26}$$

The ratio of the energy stored outside to the energy stored inside is

$$\frac{E_0}{E_i} = \frac{\dfrac{1}{b}-\dfrac{1}{(a^2-b^2)^{\frac{1}{2}}}\tan^{-1}\dfrac{(a^2-b^2)^{\frac{1}{2}}}{b}}{\dfrac{b}{a^2}-\dfrac{1}{(a^2-b^2)^{\frac{1}{2}}}\tan^{-1}\dfrac{(a^2-b^2)^{\frac{1}{2}}}{b}}. \tag{27}$$

This ratio is one measure of efficiency of a coil for intermittent operation since it indicates the relative values of useless to useful stored energy. In Fig. 3 $E_0/E_i$ is plotted as a function of the ratio $b/a$. Figure 4 shows the variation of $E_0$ as the ellipsoid parameters vary. In this figure, $6\mu E_0/(a^3B_0^2)$ is plotted as a function of $b/a$.

## VI. SPECIAL CASES OF THE ELLIPSOID

### a. The Infinite Solenoid

This familiar configuration appears when the semi-axis $b$ of the ellipsoid is allowed to become infinite. The potential $V$ outside, the external fields, and the ratio of external to internal stored energy all tend to zero. From (23) the current per unit length in the wall becomes $B_0/4\pi\mu$.

### b. The Sphere

When $b=a$, the ellipsoid becomes a sphere. From (15), (16), and (17) the external potential

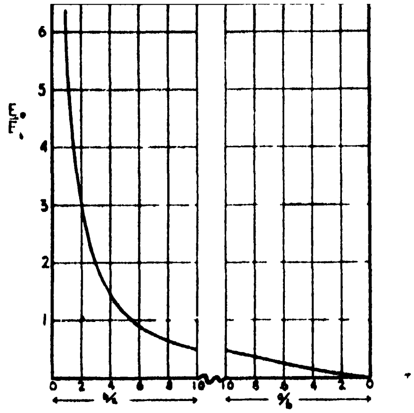


FIG. 3. Ratio of external to internal energy storage for ellipsoidal coils.

and fields are

$$V = \frac{B_0 z a^3}{2(\rho^2+z^2)^{3/2}}, \tag{28}$$

$$B_r = \frac{3B_0 z\rho a^3}{2(\rho^2+z^2)^{5/2}}, \tag{29}$$

$$B_z = \frac{B_0 a^3(2z^2-\rho^2)}{2(\rho^2+z^2)^{5/2}}. \tag{30}$$

The uniform field in the sphere is produced by a current whose density, from (21), is

$$I_\phi = \frac{3B_0\rho}{8\pi\mu a}. \tag{31}$$

From (25) the external stored energy is

$$E_0 = \frac{a^3 B_0^2}{12\mu} \tag{32}$$

and the ratio $E_0/E_i$ is given by (27)

$$\frac{E_0}{E_i} = \tfrac{1}{2}. \tag{33}$$

Spherical coils of this type have been discussed and applied to mass spectrometer design by Hipple.[5]

### c. The Plane Coil

If $b$ is reduced to zero the ellipsoidal coil becomes a flat pancake. The external potential becomes from (15):

$$V = \frac{2aB_0 z}{\pi}\left(\frac{1}{\sqrt{s}} - \frac{1}{a}\tan^{-1}\frac{a}{\sqrt{s}}\right). \tag{34}$$

From (16) and (17) the external fields are

$$B_\rho = \frac{2B_0 z\rho a^3}{\pi(2s-\rho^2-z^2+a^2)(s+a^2)\sqrt{s}}. \tag{35}$$

$$B_z = \frac{2aB_0}{\pi\sqrt{s}}\left(\frac{z^2a^2}{s(2s-\rho^2-z^2+a^2)} - 1 + \frac{\sqrt{s}}{a}\tan^{-1}\frac{a}{\sqrt{s}}\right). \tag{36}$$

In the neighborhood of this coil where $\rho$ is materially less than $a$, and $z$ is small,

$$B_\rho = \pm\frac{2B_0\rho}{\pi(a^2-\rho^2)^{\frac{1}{2}}} \text{ approx.} \tag{37}$$

the sign depending on whether $z$ is positive or negative, and

$$B_z = B_0 \text{ approx.} \tag{38}$$

From (21) the current density is

$$I_\phi = \frac{B_0\rho}{\pi^2\mu(a^2-\rho^2)^{\frac{1}{2}}}. \tag{39}$$

(The value deduced from Eq. (21) by letting $b$ tend to zero must be multiplied by a factor of two to obtain (39) since when $b=0$ the upper and lower surfaces of the ellipsoid become coincident.) From (25) the external stored energy is

$$E_0 = \frac{a^3B_0^2}{3\pi\mu}, \tag{40}$$

and the internal energy is, of course, zero.

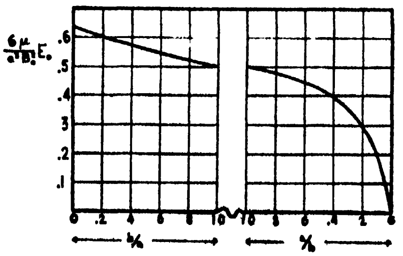


FIG. 4. External stored energy for ellipsoidal coils.

---

[5] J. A. Hipple, Phys. Rev. 55, 597 (1939).

## VII. COMPARISON OF IRON CORE AND AIR CORE FIELD COILS

As a basis on which to compare iron core and air core coils for practical applications, we shall evaluate the number of ampere turns required in the two cases to provide a uniform field $B_0$ within a given region.

We assume the ellipsoidal coil to be made up of $N$ turns, each turn carrying a current $I_0$ amperes. If the number of turns between $\rho$ and $\rho+d\rho$ is $n(\rho)d\rho$ then

$$I_0\int_{-a}^{+a} n(\rho)d\rho=\int_{-a}^{+a} I(\rho)d\rho=NI_0. \qquad (41)$$

We substitute for $I(\rho)$ from Eq. (22) and perform the integration to obtain for the total number of ampere turns:

$$NI_0=\frac{B_0 b}{2\pi\mu}\left\{\frac{a^2-b^2}{b^2-\dfrac{ba^2}{(a^2-b^2)^{\frac{1}{2}}}\tan^{-1}\dfrac{(a^2-b^2)^{\frac{1}{2}}}{b}}\right\}. \qquad (42)$$

The number of ampere turns $N'I'$ required in an iron core coil for generation of a field $B_0$ webers per sq.m in an air gap of length $l$ meters is

$$N'I'=\frac{B_0 l}{4\pi\mu}=8.0\times 10^5 B_0 l \qquad (43)$$

provided the air gap is short compared with the other dimensions of the region in which the field is to be maintained. If the air gap becomes comparable with the other dimensions the number of ampere turns will be materially higher than the value given by (43), since the fields begin to suffer in strength and uniformity as edge effects become more important.

A comparison between Eqs. (42) and (43) shows that for small air gaps the iron core has marked advantages. For an air gap one-tenth of the pole-piece diameter, the iron core magnet requires only about one-eighth of the ampere turns needed by the corresponding ellipsoidal air core coil for which $b/a=0.1$. When the air gap becomes about half the pole-piece diameter, the magnets become comparable in ampere turn requirements, and for larger air gaps the ampere turn requirements remain comparable. Evidently the mechanical simplicity and compactness of the air core coil merit consideration in any case where the air gap in the corresponding iron core magnet would be greater than half the pole-piece diameter.

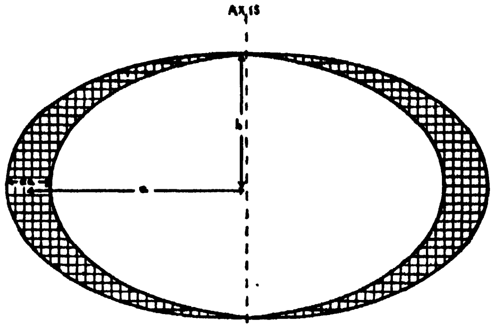


FIG. 5. Cross section through ellipsoidal coil for production of uniform field.

## VIII. POWER CONSUMPTION IN ELLIPSOIDAL D.C. FIELD COILS

The distribution of turns in the wall of the ellipsoid can be expressed rather simply using Eqs. (22), (41), and (42). The number of turns between $\rho$ and $\rho+d\rho$ is

$$n(\rho)d\rho=\frac{N\rho d\rho}{2a(a^2-\rho^2)^{\frac{1}{2}}}. \qquad (44)$$

Similarly, the number of turns between $z$ and $z+dz$ is

$$n(z)dz=\frac{Ndz}{2b}. \qquad (45)$$

The total length of wire used in the winding will be

$$\int_{-a}^{+a} 2\pi\rho n(\rho)d\rho=\frac{\pi^2 aN}{2}\ \text{meters}. \qquad (46)$$

If the coil is wound with copper wire having a cross-sectional area of $s$ sq.m. the total resistance of the winding will be

$$R_0=1.7\times 10^{-8}\frac{\pi^2 aN}{2s}\ \text{ohms}. \qquad (47)$$

From (42) and (47) the power consumption in the coil will be

$$I_0{}^2R_0=\frac{1.9\times 10^5 ab^2B_0{}^2}{Ns}\times\left\{\frac{a^2-b^2}{b^2-\dfrac{ba^2}{(a^2-b^2)^{\frac{1}{2}}}\tan^{-1}\dfrac{(a^2-b^2)^{\frac{1}{2}}}{b}}\right\}^2\ \text{watts}. \qquad (48)$$

If the thickness of the coil in the $z=0$ plane is $\alpha a$ (from (23) or (45) $\alpha a$ will also be the thickness in any plane parallel to the $z=0$ plane) then from (45)

$$Ns=2\alpha ab \tag{49}$$

and the power consumption may be written

$$I_0{}^2R_0=\frac{9.5\times 10^4 bB_0{}^2}{\alpha}\times\left\{\frac{a^2-b^2}{b^2-\dfrac{ba^2}{(a^2-b^2)^{\frac{1}{2}}}\tan^{-1}\dfrac{(a^2-b^2)^{\frac{1}{2}}}{b}}\right\}^2 \text{ watts.} \tag{50}$$

For $b/a=0.6$, 1.0, and infinity, the quantity in the bracket will have the values 1.9, 1.5, and 1.0, respectively.

Figure 5 is a sketch of the cross section through an ellipsoidal coil for which $b/a=0.6$ and $\alpha=0.2$.

## VIII. ACKNOWLEDGMENTS

The author is indebted to Professor F. W. Grover for a considerable amount of information on techniques for dealing with the circular loop problem and with elliptic integrals. Professor Grover's study of the fields around a circular loop will appear in a forthcoming paper in the *Proceedings of the Institute of Radio Engineers.* The tables of magnetic fields around a circular loop represent a long period of patient and painstaking calculation on the part of Miss M. Kay.

---

# Design of an Air Core Synchrotron*

JOHN P. BLEWETT**
*East Patchogue, Long Island, New York*

(Received June 17, 1947)

The analysis of the preceding paper is applied to the problem of designing a synchrotron in which the magnetic fields are produced by air core coils. The air core synchrotron is shown to be practical and to present certain advantages.

## I. INTRODUCTION

In the preceding paper an analysis has been presented of the problems, first, of the magnetic fields caused by current flowing in a circular loop, and second, of the coil configurations necessary to produce uniform fields in regions of various geometries. The magnetic field requirements of the betatron and the synchrotron will now be considered in the light of this analysis.

Air core magnets present three major advantages. First, they free the magnet designer from the upper limit of 1.8 webers per square meter (18,000 gausses) or so associated with iron cores because of the saturation characteristics of iron. Second, they do not introduce the random local inhomogeneities in field which are found in the fields of iron core magnets as a result of defects in fabrication or spottiness in previous treatment. This drawback of iron cores may be a serious one in the betatron or the synchrotron in which injection takes place just after the field has passed through zero. If all parts of the field do not pass through zero simultaneously it may be difficult to initiate a beam, particularly in the higher energy machines. The third advantage of air core coils is the obvious one that a large and costly iron structure is avoided. This is particularly important in betatrons and synchrotrons where the iron must be laminated, and the core construction is an extremely tedious and expensive proposition.

The problems which arise when a switch is made from iron to air core are mainly associated with the choice of the correct geometry for the

---

* Most of the work described in this paper was supported by the Office of Naval Research under Contract N5-ORI-178.

** Formerly Research Laboratory, General Electric Company. Now at Brookhaven National Laboratory, Brookhaven, Long Island.

current carrying conductors. Possible solutions to these problems are discussed in Sections II, III, IVa, and IVb. Other problems associated with air core coils such as cooling, auxiliary circuits, mechanical support, and so forth do not seem unduly serious. Typical solutions of several such problems are presented in Sections IVc, IVd, and IVe.

## II. AIR CORE GUIDE FIELDS FOR BETATRONS OR SYNCHROTRONS

The primary economic factor in the design of a betatron or synchrotron, aside from the magnet itself, is the capacitor bank in which the energy to appear in the magnetic field must be stored during the inactive periods. It is therefore important that as much of the field generated as possible shall be actually useful. An iron core magnet designed to produce a field of a particular type generally stores two or three times as much energy as is required for fields which actually deflect particles. This is simply because the shape of the pole piece which produces the desired field is generally not the same as the shape of the vacuum envelope in which the field is utilized. In addition wasted energy is always stored in leakage fluxes. The air core magnet can be designed to fit the vacuum envelope. It must, however, have a finite thickness to permit it to dissipate power. And it must store unused energy in return fluxes.

For toroidal regions such as those enclosed by the vacuum envelopes of betatrons and synchrotrons, air cores and iron cores are about equally economical in stored energy. The problem of generation of uniform fields in a toroidal region has not yet been solved analytically so far as the author is aware. An approximate attack is suggested, however, by the infinite cylinder case treated in Section III of the preceding paper. If the cylinder is bent in the $x, y$ plane on a radius of curvature materially greater than its own radius, the external distribution of energy will not be too violently disturbed. It follows from Eq. (11) of the preceding paper that three-quarters of the energy in the external field of the cylinder lies between the cylinder wall (radius $R$) and a cylinder of radius $2R$. This will probably still be approximately true when the cylinder has been bent into a torus. It is therefore a reasonable assumption that the energy stored in the external fields of the torus is approximately the same as the energy stored within the torus wall, just as was true in the case of the infinite cylinder. In other words, in setting up a uniform paraxial field within a toroidal region we shall store approximately twice the energy we use because of the energy stored in the external return flux. If now we arbitrarily increase the factor of two to a factor of three to take care of the excess energy stored because of the necessarily finite thickness of the coil, we shall have a design factor not far from that actually realizable.

If a roughly uniform field $B_0$ webers per sq. m is set up within a toroidal region whose cross-sectional area is $\pi r^2$ sq. m and whose mean radius is $R$ m, the energy stored in this region will be

$$E_i = \frac{\pi R r^2 B_0^2}{4\mu} \text{ joules,} \tag{1}$$

where $\mu$, the permeability, $=10^{-7}$ henry per meter for free space. To obtain the total energy stored in all fields we multiply this quantity by our factor of three to obtain

$$E = \frac{3\pi R r^2 B_0^2}{4\mu} = 2.4\times 10^7 R r^2 B_0^2 \text{ joules.} \tag{2}$$

This equation can be rewritten in terms of the energy $W$ ($=mc^2/e$ electron volts) of a particle bent in a circle of radius $R$ in a field $B_0$ through the relation

$$W = B_0 c^2 R/v \text{ electron volts,} \tag{3}$$

where $v$ is the particle velocity in meters per sec., $c$, the velocity of light, $=3\times 10^8$ m per sec. From (2) and (3)

$$E = 2.7\times 10^{-10}\frac{r^2 W^2}{R}\frac{v^2}{c^2}. \tag{4}$$

If, for example, $R$ is one meter, the vacuum envelope is a torus of three-inch tubing so $r=0.038$ m and $W$ is 100 Mev, $v$ will be approximately equal to $c$ if the particles are electrons, and $E$ will be about 4000 joules. This much energy could be stored at 5000 volts in sixteen 20-microfarad capacitors.

If this device is to be operated continuously

TABLE I.

| $W$(Mev) | $R$(m) | Number of capacitors for 5-Mev betatron | Number of capacitors for guide field |
|---|---|---|---|
| 100 | 0.5 | 7 | 30 |
| 500 | 2.5 | 35 | 150 |
| 1000 | 5.0 | 70 | 300 |

at a frequency $\omega/(2\pi)$ the power required may be deduced if an estimate can be made of the $Q$ of the coil. From the definition of $Q$,

$$\text{power required} = \omega E/Q \text{ watts.} \tag{5}$$

At 60 cycles, $Q$'s higher than 5 or 6 are hard to achieve in air core coils. For a $Q$ of 5 and the coil discussed above the power required would be about 300 kilowatts. This is rather high and might justify reconsideration of the higher $Q$'s obtainable in iron core coils. But if, for example, the machine is to be used for cloud-chamber investigations, it need be fired only once or twice a minute. In this case the power consumption drops to the negligible range below 200 watts.

The actual design of guide-field coils will be treated in more detail in Section IV.

## III. AIR CORE FLUX COILS FOR BETATRON ACCELERATION

In iron core machines the energy stored in fluxes used for betatron acceleration can be negligible if air gaps are kept out of the flux paths. This is because the energy stored for a given flux density is inversely proportional to the permeability of the medium. The air core betatron is therefore at a disadvantage so far as flux coils are concerned. To obtain order of magnitude figures for air core flux coil energy requirements we shall now consider an idealized case.

Efficient energy storage will be achieved if we keep the flux density uniform within the orbit. This could be done with a pancake coil of the sort discussed in Section VIc of the preceding paper. The flux through such a coil of radius $R$ will be

$$\phi = \pi R^2 B_0 \text{ webers.} \tag{6}$$

But from the betatron condition

$$\phi = 2\pi R^2 B_1 \text{ where } B_1 \text{ is the guide field at the orbit}$$

$$= \frac{2\pi R}{c}\frac{v}{c} W \text{ webers} \tag{7}$$

(compare Eq. (3) above). But from Eq. (38) of the preceding paper the energy stored by the flux coil is

$$E_f = \frac{R^3 B_0^2}{3\pi\mu} \text{ joules.} \tag{8}$$

From (6), (7), and (8)

$$E_f = \frac{4R}{3\pi\mu}\frac{v^2}{c^2}W^2 = 4.7\times 10^{-11} R W^2 \frac{v^2}{c^2} \text{ joules.} \tag{9}$$

If $W = 10$ Mev, $R = 0.5$ m, and $v$ is approximately equal to $c$, then $E_f = 2400$ joules. If $W = 100$ Mev, $R = 1$ m, and $v$ is approximately equal to $c$, $E_f$ will be 470,000 joules. In the former case the energy could be stored in ten 5000 volt, 20 microfarad capacitors, while in the latter case 2000 such capacitors would be required. Which, if either, of these quantities is reasonable depends, of course, on the designer's purse and viewpoint. It is evident in any case that air core betatrons for higher energies than 20 Mev will begin to be expensive.

In view of the idealized geometry considered, the figures derived from Eq. (9) are merely lower limits. It would be impractical to extend the flux coil all the way to the orbit and it would, moreover, be impractical to use a pancake coil since its external fields which would exist at the orbit are strongly radially dependent and would interfere with the guide field. A more practical design will be presented in the next section. For the practical flux coil, the factor of 4.7 in Eq. (9) should be replaced by a factor of about 14 to give

$$E_f = 1.4\times 10^{-10} R W^2 \frac{v^2}{c^2} \text{ joules.} \tag{10}$$

## IV. DESIGN OF AN AIR CORE SYNCHROTRON

In the introduction we have pointed out the advantages of air core machines. In Section II air core guide fields were shown to be economically comparable with iron core guide fields, particularly for intermittent operation, and in Section III it was shown that betatron acceleration up to 10 Mev with air core flux coils involves not unreasonable figures for energy storage. Since betatron acceleration up to four or five Mev is quite sufficient to initiate the usual sequence of events in the synchrotron, the betatron coils of an air core synchrotron could just as well be air core. Introduction of an iron core for the betatron flux would result in rather a tricky problem of synchronization between air core fields and iron core fluxes. It would therefore appear from what has gone before that a high energy air core synchrotron is quite practical. If we speak in terms of 5000-volt 20-microfarad capacitors and of a three-inch diameter vacuum envelope, the results of Eqs. (4) and (10) are as given in Table I for a few typical cases.

We now proceed to a more detailed consideration of design.

### a. Guide-Field Coils

The guide field must decrease with radius at a rate not faster than $1/\rho$ in order that the electron motion shall be stable both in $\rho$ and in $z$. For most designs this means a variation across the vacuum envelope of not more than about ten percent, and the guide-field coils will not differ much from those required to produce a uniform field. As was mentioned in Section II, this problem has not been solved analytically, but a measure of similarity exists between the torus and the infinite cylinder. We should, therefore, begin with a coil system suggested by the cylinder case, such as that shown in cross section in Fig. 1, and, by numerical work based on the tables in the preceding paper or by measurements on models, make the small changes which will result in the desired variation of field with radius.

### b. Flux coils

The flux coils will be required to build up a flux at a rate governed by the rate of rise of the guide field to the point at which the electron energy is about 5 Mev. Now the radiofrequency signal should be turned on and synchrotron acceleration initiated. The guide field should continue to rise, but the betatron flux should cease rising since further increases will result merely in useless storages of energy. In iron core machines this object is accomplished by saturation of the flux core. In an air core machine different methods must be used. One possible method is described in Section c below.

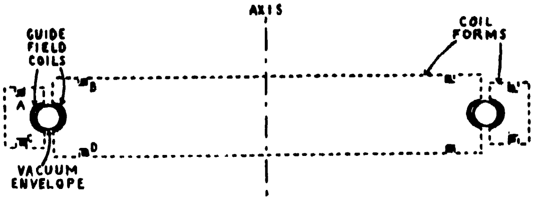


FIG. 1. Cross section through air core synchrotron.

Since the time variations of field and flux are to be different, it follows that no appreciable field from the flux coil can be permitted to exist in the neighborhood of the orbit. On the other hand, in order that we do not store large amounts of energy needlessly in the flux, it is desirable that the flux coils couple as closely as possible with the orbit. These apparently mutually exclusive requirements can be met by the compromise to be described below.

In the case of the infinite cylinder discussed in Section III of the preceding paper, had we set up currents $I_x$ in the walls proportional to the absolute value of $y$ instead of $y$, the internal field would have been zero. If, now, a wire parallel to the cylinder axis were located somewhere within the cylinder unity coupling would exist between it and the currents in the cylinder walls since no magnetic flux interposes between wire and wall. Presumably an analogous case exists for the toroid. The exact distribution of azimuthal current in a toroid wall, which would give zero internal field, has not yet been worked out. A fairly good approximation can easily be set up, however, using only four wires appropriately located in the torus wall. The four wires $A$, $B$, $C$, and $D$ of Fig. 1 lie in one possible configuration. The choice of positions for these four wires is made after a consideration of the field distributions around a circular loop discussed in Section II of the preceding paper and shown graphically in Figs. 1 and 2 of that paper. Wire $A$ is located so that the maximum in its field pattern lies at the orbit. The position of wire $B$ is such that the minimum in its field pattern lies at the orbit, and the relative currents in the two wires are so adjusted that the minimum and maximum are equal in height. In this way the axial field at the orbit and its first derivative with respect to radius are reduced to zero. The selection of parameters may be expedited by using the graph in Fig. 2 of the preceding paper. In applying this method it will become evident that the

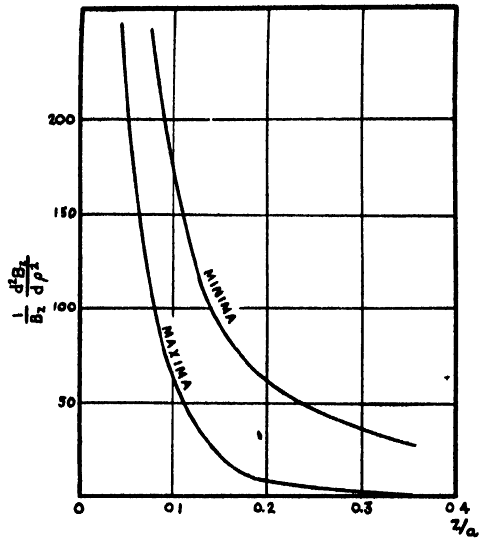


FIG. 2. Second derivatives at maxima and minima in fields of circular loop such that these extrema lie at $\rho=1$.

choice of location and size of the coils is not unique. For each coil, loci of the form $z=f(\rho)$ exist along which cancellation can be achieved by appropriate choices of current ratios. This extra degree of freedom can be utilized to permit cancellation of second derivatives. To facilitate this procedure, the ratios of second derivatives of $B_z$ with respect to $\rho$, to the values of $B_z$, have been plotted in Fig. 2 for the cases where the extrema of the field distributions lie at $\rho=1$. When the value of $z/a$ for one coil has been chosen, the $z/a$ for the second coil will be that for which the ordinates of the curves of Fig. 2 have equal values. Given $z/a$ for each coil, the correct value of $\rho/a$ may be found from Fig. 2 of the preceding paper. The results of this rather complex procedure are summarized in the graphs of Fig. 3 which give complete design data for pairs of loops which give cancellation of the field, its first and second derivatives, at the orbit. For the case where the orbit is located at $\rho=1$ the radius $a_0$ and height $z_0$ of the outer coil and the radius $a_i$ of the inner coil are plotted as functions of the height $z_i$ of the inner coil. Also included in Fig. 3 is the ratio of currents in the two coils as a function of $z_i$. This graph indicates that for the geometry shown in Fig. 1 the ratio of ampere turns in the inner coil to ampere turns in the outer coil is about 4:1.

After cancellation of axial fields and derivatives thereof has been accomplished, addition of the two coils $C$ and $D$, symmetrically located below the plane of the orbit, will give cancellation of radial fields.

It is unnecessary to add further refinements to provide cancellation of fields adjacent above and below the orbit. If $B_z$, and its derivatives with respect to radius, and $B_\rho$ are zero in the plane of the orbit, the fact that the curl and divergence of the field are zero insures the disappearance of the fields at all points in the neighborhood of the orbit.

A further improvement in cancellation and a closer approximation to the ideal case would result from the use of six or more coils so adjusted that still higher derivatives of the field would disappear. This rather formidable problem has not yet been worked out.

Experimentally it has been found that the coupling coefficient between the orbit and a set of coils in about the configuration of Fig. 1 is between 0.8 and 0.9. In other words, the energy gained per revolution by an electron in this system will be between 80 and 90 percent of the voltage per turn applied to the flux coil. This result was obtained with coils about 20 cm in

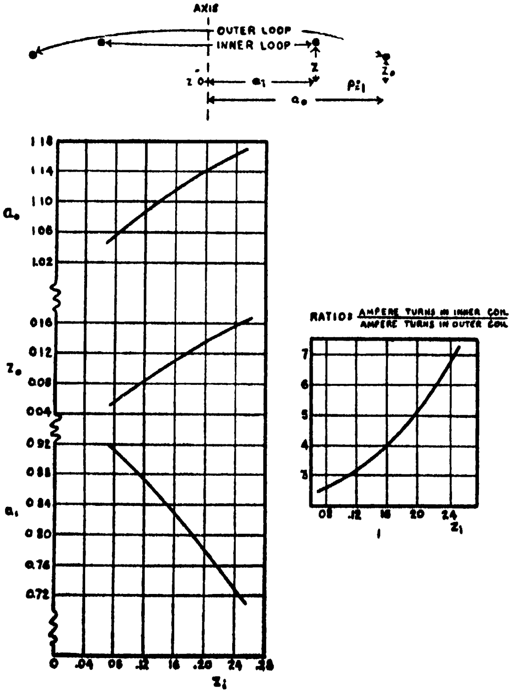


FIG. 3. Design parameters for matching second derivatives.

radius and having cross sections of about one sq. cm. The coupling coefficient might be materially lower for coils of smaller cross section.

### c. Construction of Field and Flux Coils

Two important problems must be kept in mind during the actual construction of the field and flux coils. The first is associated with the forces between conductors. In a high energy machine these forces may be of the order of several tons per meter of coil. The forces will be attractive between conductors passing currents in the same direction and repulsive between conductors passing currents in opposite directions. These forces, fortunately, are such that the coils will tend to be held in place, rather than explode, in a coil form of the convenient shape indicated by dotted lines in Fig. 1. During several months of model experiments on a small 100-Mev design for intermittent operation no difficulty was experienced with these forces.

The second problem lies in the discontinuities at points where leads enter or leave the coils. Theoretically, if entering and return leads lie very close together, no discontinuity will appear, but in practice some space must be left for insulation and detectable inhomogeneities in field do appear. This problem was solved in model experiments by making each coil several coils in parallel. The leads to the several coils were brought out at equally spaced points around the coil circumference. The entering and return leads to each coil were then carried close together and parallel to the coil axis a distance of about one coil radius. They were then bent inward and carried together to the coil axis. All connections were made on the coil axis. Measurements good to 0.1 percent on fields in the neighborhood of the orbit failed to show any measureable inhomogeneity when each coil had been made six coils in parallel. When coils are paralleled thus it is, of course, necessary to interweave the windings in such a way as to maintain uniform current density through the coil and avoid circulation of parasitic currents.

Table II.

| Phase of sine wave: | 5° | 10° | 15° | 20° | 25° |
|---|---|---|---|---|---|
| Deviation of sine from linear expression (%): | 0.13 | 0.51 | 1.14 | 2.03 | 3.18 |

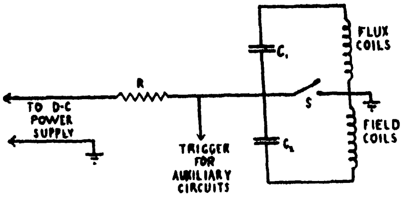


Fig. 4. Circuit for air core synchrotron.

### d. Auxiliary Circuits and Equipment

The requirement that the flux should cease rising after the electron velocity is approximately that of light (compare Section IVb above) can be met by making the flux coils part of a circuit whose resonant frequency is several times that of the circuit which includes the field coils. When a capacity $C$ charged to a voltage $V$ is discharged into an inductance $L$, the current is given by

$$i = V(C/L)^{\frac{1}{2}} \sin(t/(LC)^{\frac{1}{2}}). \tag{11}$$

For $t$ small, $i = Vt/L$ approximately. As $t$ increases the expression (11) deviates from linearity with $t$ in accordance with Table II. If these figures relate to the higher frequency flux wave, the lower frequency field wave will be essentially linear, and the table gives the percentage error from the betatron condition. The equilibrium orbit will change to a new radius, and the ratio of the change in radius to the original radius can be shown to be essentially the same as the deviation given in the above table.

Since the vacuum envelope must certainly be large enough that a variation in orbit radius of two or three percent can be tolerated, betatron acceleration will be satisfactory up to 20° or 25° of the flux wave. At this point the radiofrequency acceleration must be initiated. We shall now store energy uselessly in the flux coils until 90° in the flux wave, an inelegant feature of this method which is partly counterbalanced by the method's simplicity in execution.

As the flux wave falls off past 90° we shall have a negative betatron effect, but the stability of the synchrotron orbit has often been shown to be sufficient to offset this provided the radiofrequency signal level is sufficiently high.

A circuit for generating the desired sequence of events is shown in Fig. 4. The field and flux capacitor banks, $C_2$ and $C_1$, are charged as indicated from the same power supply. They are then discharged simultaneously by the switch $S$ which can be a large circuit breaker or a bank of ignitrons. At the moment of closing, the potential of the left-hand contact of the switch drops discontinuously to zero. This sharp wave front provides a good trigger for timing injection and radiofrequency turn on and off.

Design data for timing and injection circuits, vacuum envelopes, cooling systems, electron guns, and radiofrequency systems are available in profusion and will not be discussed here.

### e. Summary of Test Results

A program of construction of a 100-Mev air core synchrotron was initiated by the author in the Research Laboratory of the General Electric Company, but his departure from that laboratory prevented him from carrying the program to completion. Numerous model measurements were made on field and flux coils and preliminary circuit tests were performed, but the construction of circuits and electron gun did not reach the point where electrons could be accelerated.

The machine had an orbit radius of 7 inches and was designed for once per minute operation. The structure was generally as indicated above except for the guide field coils which were simply coaxial 44 turn solenoids about 3 inches long. The outer solenoid was divided into two halves in a plane normal to its axis, and it was found that a paraxial separation of about an eighth of an inch of these two half-coils gave a field in the neighborhood of the orbit which fell off approximately as the inverse half-power of radius. This coil design was not as economical as the one indicated in Fig. 1 should be, since considerable amounts of energy were stored in regions above and below the vacuum envelope.

The coils were wound on Bakelite forms generally similar to those shown in Fig. 1.

The vacuum envelope was a toroid of one-inch glass tubing. The electron gun was brought in from the top; other associated equipment was conventional in design.

Once a minute operation solved all cooling problems. It did, however, introduce a difficulty which might well be eliminated in any future machines of this type. Experience with the 70-Mev iron core synchrotron in the General Electric Research Laboratory has indicated that the serious problems in synchrotron operation are all betatron problems; if successful betatron operation can be achieved, further acceleration by radiofrequency fields is very easy. It would seem, therefore, that even though a machine is to be used for intermittent high energy operation it should still be capable of continuous low energy betatron operation. This will speed up very materially the initial tests when a number of parameters are varied in attempts to find a beam. Given a beam, further adjustments to maximize beam intensity or to carry it to higher energy are relatively simple and can be carried on at a once per minute rate without reducing the operator to a state of impatient exasperation.

No special means were included for bringing out the beam. It was hoped eventually that the beam could be deflected into the space above the orbit. It seems on paper to be much easier to move or at least to tilt the beam axially, than it is to expand it in a controlled fashion. In iron core machines, of course, the pole pieces are in the way. In an air core machine a weak auxiliary field might be sufficient to bring a sharp beam into the clear region above the coils.

## V. ACKNOWLEDGMENTS

The author is indebted to Mr. L. Redmond for his invaluable assistance in the performance of the experiments discussed above. Mr. A. M. Gurewitsch, Dr. R. V. Langmuir, Dr. E. E. Charlton, Dr. G. Ragan, and Dr. J. L. Lawson and many others of the author's former associates in the General Electric Research Laboratory have contributed materially in assistance and suggestions. Finally, it is a pleasure to acknowledge the continued advice and support of Dr. A. W. Hull.

# A Sound Velocity Method for Determining the Compressibility of Finely Divided Substances

R. J. URICK
*Naval Research Laboratory, Washington, D. C.*

(Received June 23, 1947)

A method is presented whereby the adiabatic compressibility of a finely divided material can be found from sound velocity and density measurements of a suspension of the particles in a liquid. The method is based on the assumption that the velocity of sound in a suspension is the same as it would be in an ideal solution of the two substances. This is verified experimentally by measurements of sound velocity in kaolin-water suspensions and xylene-water emulsions, and the method is illustrated by the determination of compressibility of the oil droplets in an oil emulsion and of the blood corpuscles in horse blood. Considerable accuracy is attainable for materials which are not too incompressible compared to the suspending liquid.

## INTRODUCTION

THE ultrasonic interferometer has been employed for many years for the determination of the adiabatic compressibility of liquids. This instrument serves to measure the wavelength of sound waves in the liquid, and thereby the velocity of sound, and a simple calculation yields the compressibility once the density of the liquid is known. Extensive tables of compressibility of pure liquids are available.[1]

In 1930, K. H. Herzfeld[2] suggested that this method might be extended to the study of solid substances by grinding them, suspending the particles in a liquid, and measuring the velocity of sound in the suspension. He derived an expression for the adiabatic compressibility of the solid in terms of the velocity, density of solid and liquid, and the compressibility of the liquid. This was based on extensive theoretical considerations of the effect of scattering by numerous small rigid spherical particles in the sound field. The result was applied for this purpose by Randall[3] a short time later, but extremely small, or even negative, compressibilities were found.

The failure of Herzfeld's formula to find experimental verification seems to have been followed by a long period of neglect in the application of the ultrasonic interferometer to substances other than homogeneous liquids and gases.

In the present paper a simple expression will be derived for the velocity of sound in a suspension based on the assumption that the mixture behaves the same as would an ideal solution of the two substances. The result will be verified by some data on kaolin-in-water suspensions and on some emulsions of xylene and water. These considerations permit a determination of the adiabatic compressibility of the particles in a suspension from measurements of sound velocity and density, and this method will be illustrated by the computation of the compressibility of (1) the oil particles in some oil-in-water emulsions and (2) blood corpuscles in horse blood.

## SOUND VELOCITY IN IDEAL MIXTURES

In approaching the subject of the velocity of sound in a suspension or emulsion of one substance in another, it is simplest to consider what takes place in a homogeneous ideal mixture of the two materials. In so doing, we must make the assumption that the suspended particles are infinitesimally small compared to the wavelength of the sound, and that accordingly the effects of scattering on the sound velocity may be neglected. This is seen to be not too unreasonable an assumption when it is realized that at a frequency of one megacycle the size of the particles in a course dispersion is less than a hundredth of the wave-length of sound in water. Making this assumption, we may proceed to find the sound velocity in the same manner as for an ideal solution of one substance in the

---

[1] L. Bergman, *Der Ultraschall* (Edwards Bros., Ann Arbor, Michigan, 1942), pp. 175–181.

[2] K. H. Herzfeld, "The propagation of sound in suspensions," Phil. Mag. 9, 752–68 (1930).

[3] C. R. Randall, "Ultrasonic measurements of the compressibility of solutions and of solid particles in suspension," *Bur. Stand. J. Research* 8, 79–99 (1932).

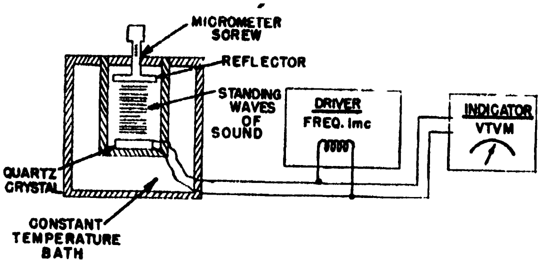


FIG. 1. Diagram of 1-megacycle acoustic interferometer.

other.[4] In this case both the density and the compressibility of the mixture are additive properties of the corresponding quantities for the two materials, depending upon the proportional amount of each substance in the mixture. Thus, $\rho_0=\rho_2\beta+\rho_1(1-\beta)$ and $\kappa_0=\kappa_2\beta+\kappa_1(1-\beta)$, where $\rho_0$, $\kappa_0$ are the density and compressibility of the suspension, $\rho_2$, $\kappa_2$ are the density and compressibility of the suspended particles, $\rho_1$, $\kappa_1$ are the density and compressibility of the suspending substance, $\beta$ is the volume percentage of particles. The subscripts 0, 1, and 2 refer to the suspension, suspending liquid, and the suspended particles, respectively. Hence, the velocity of sound in the mixture is

$$V_0=\frac{1}{(\rho_0\kappa_0)^{\frac{1}{2}}}=\left[\frac{1}{[\rho_2\beta+\rho_1(1-\beta)][\kappa_2\beta+\kappa_1(1-\beta)]}\right]^{\frac{1}{2}}.$$

After dividing by $V_1^2=1/\rho_1\kappa_1$, this expression may be simplified in form by writing, $\gamma=V_1/V_0$, the ratio of the velocity of the suspending liquid to that of the suspension, $a=(\rho_2-\rho_1)/\rho_1$, $b=(\kappa_2-\kappa_1)/\kappa_1$, and we may then obtain

$$\gamma^2=(1+a\beta)(1+b\beta). \qquad (1)$$

Thus the velocity ratio $\gamma$ is a parabolic function of the concentration $\beta$ and will therefore have a maximum or a minimum at some particular concentration $\beta_m$. This concentration,

$$\beta_m=-(a+b)/2ab$$

will be physically realizable, of course, only if it happens to lie between 0 and plus 1. It will be seen later that a minimum of velocity was found for mixtures of kaolin and water.

### THE INTERFEROMETER

The ultrasonic interferometer consists of a piezoelectric crystal mounted in an enclosed chamber. At the opposite end is a reflecting plate capable of being moved parallel to the direction of propagation of the sound by a micrometer screw. The crystal is driven through a loosely coupled coil by a source of constant frequency, and the voltage across the crystal is indicated by a meter. This system is shown schematically in Fig. 1. As the micrometer screw is moved outward or inward, the meter is found to exhibit regular deflections at intervals of a half wave-length, because of the reaction of the standing wave system upon the driving crystal. Thus, the micrometer serves to measure the half wave-length, and since the frequency is known, the velocity of sound in the chamber is readily found. The interferometer chamber is surrounded by a bath whose temperature can be kept constant to 0.01° or so. It is estimated that the interferometer enabled velocity measurements to be made with an average error of 1 part in 3000 or 0.03 percent.

The driving frequency was kept constant at 1 megacycle. At this frequency the wave-length of wound in water is 1.5 mm; the particle size of the suspensions to be described was in the range 1 to 10 microns.

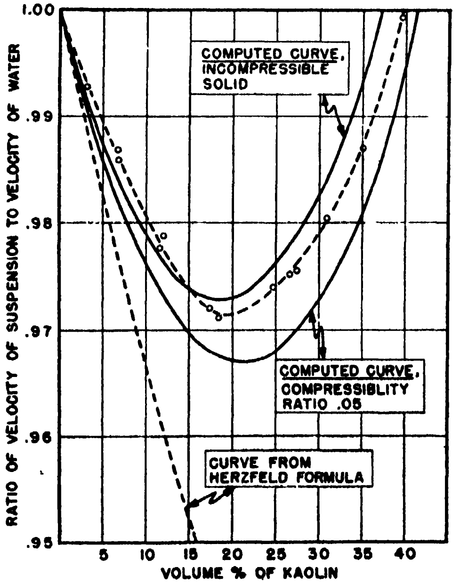


FIG. 2. Sound velocity ratio *vs*. concentration, kaolin suspensions.

---

[4] A. B. Wood, *A Textbook of Sound* (G. Bell & Sons, London, 1941), pp. 361–362.

## KAOLIN SUSPENSIONS

The validity of the additive formula for sound velocity may be demonstrated by some data on suspensions of kaolin in water. Kaolin is a complex aluminum silicate occurring in nature in a rather high degree of purity. It is a residual weathering product and a common constituent of natural clays and muds. It is an admirable material for providing solid aqueous suspensions since it disperses readily in water, does not gel, and needs no grinding to render the particles fine. By using a suitable deflocculant, such as sodium pyrophosphate, in an amount necessary to give a pH of 9–10, suspensions of kaolin with a volume concentration of 35–40 percent can be made rather readily with sufficient fluidity for use in the interferometer. Care was taken to remove the air mixed into the suspension by the dispersion process. A size analysis showed that 90 percent of the particles were between 0.4 and 5.0 microns in diameter, with a median size of 1.1 microns. The concentration of kaolin was determined by density measurements, using a density of 2.60 for kaolin.

Figure 2 shows a plot of sound velocity in such mixtures against percentage by volume of kaolin. The velocity is plotted as the ratio of velocity in the suspension to that of the liquid (water and deflocculant) in which the kaolin particles are suspended. The velocity at first falls as kaolin is added to the liquid, reaches a minimum at 20 percent kaolin by volume, then rises. The two solid curves are computed from Eq. (1) for a compressibility ratio of 0.05 and zero times the compressibility of the suspending liquid. The zero compressibility-ratio curve refers, of course, to the limiting case of an incompressible solid. Also plotted in this figure is a curve computed from Herzfeld's formula mentioned above.

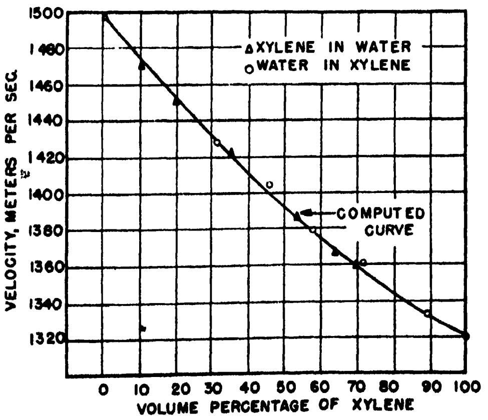


FIG. 3. Sound velocity *vs.* concentration for xylene-water emulsions.

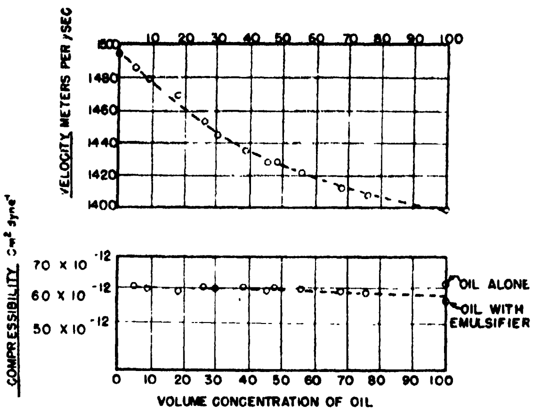


FIG. 4. Sound velocity *vs.* concentration for transformer oil-in-water emulsions.

## EMULSIONS

The applicability of the additive formula can be further illustrated by some emulsions of xylene-in-water and its inversion, water-in-xylene. A xylene-in-water emulsion consists of small separate globules of xylene suspended in a continuous volume of water; such emulsions were made by using sodium oleate as emulsifier. On the other hand, in a water-in-xylene emulsion the globules are water and the continuous liquid is xylene; these were prepared with the aid of sorbitan monooleate as emulsifier. As is common in emulsion technology, phase reversal was avoided by adding a small part of the discontinuous medium (say xylene in the xylene-in-water emulsions) to all of the other medium plus all the emulsifier, then shaking to produce an emulsion, and repeating the process for more and more of the first medium, until all of it was emulsified. The type of emulsion finally obtained was checked by the simple method of observing the miscibility of the emulsion with water and with xylene. Measurement of particle size in these emulsions with a microscope indicated a median particle size of approximately 5 microns.

Velocity measurements made over a range of concentration for the two types of emulsions are

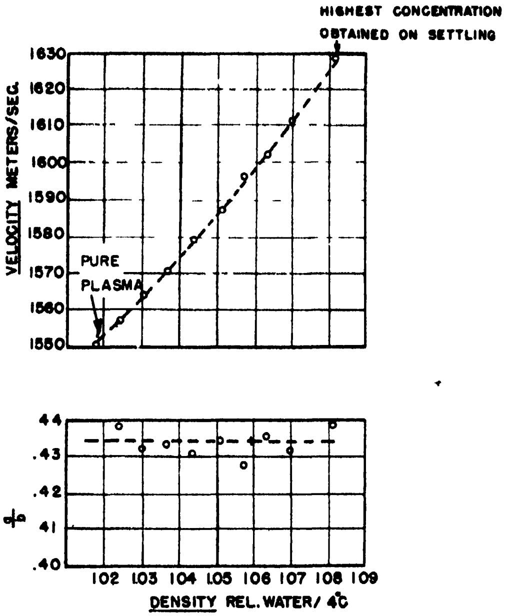


FIG. 5. Sound velocity and the ratio $a/b$ for various dilutions of horse blood with plasma.

shown in Fig. 3. It is interesting to observe that both types of emulsions have the same sound velocity at the same concentration of xylene, indicating that the sound velocity is determined by the composition of the emulsion. The measured sound velocity and density of the end members, xylene and water, enable their compressibility to be found, and Eq. (1) yields the velocity for different mixtures of the two. The velocity so computed is plotted as the solid line in this figure.

## DETERMINATION OF COMPRESSIBILITY

The additive formula provides a method for determining the compressibility of a material which occurs as, or can be rendered into, fine particles. Since the density of the particles is known, it is necessary only to suspend them in a liquid of known or measurable density and sound velocity, and measure the density and velocity of the suspension. For this purpose, we may write Eq. (1) as

$$b=\frac{\kappa_2-\kappa_1}{\kappa_1}=\frac{1}{\beta}\left(\frac{\gamma^2}{\rho}-1\right) \qquad (2)$$

where $\rho$ is the ratio of the density of the suspension to that of the suspending liquid.

The application of this formula may be illustrated by some data obtained on emulsions of transformer oil in water. These emulsions were made by adding 8 percent by weight of triethanolamine oleate to the oil as emulsifier, and agitating thoroughly with water to produce an emulsion. This particular emulsifier is very soluble in oil but only slightly soluble in water. An emulsion of high oil concentration was made in this manner, and those of lower concentration were obtained through successive dilutions with water. Occluded air was removed by placing the emulsion in a vacuum until no further air was seen to come off. Velocity and density measurements were made on eleven such emulsions of varying percentage of oil, and also on water alone and on the oil with and without emulsifier. Figure 4 shows sound velocity plotted against concentration, and in the lower part of the figure, the computed values of compressibility of the oil. These values are seen to be nearly, but not quite constant, and lie intermediate between the measured values for the oil with and without emulsifier. The slight change of compressibility with concentration in this case may represent the effect of partial leaching out of emulsifier from the oil by the successive dilutions with water.

In a similar manner there may be obtained an approximate value of compressibility of the kaolin particles in the suspensions of Fig. 2. Using the data for the three highest kaolin concentrations, $b$ may be found to be $-0.984$, $-0.972$, and $-0.977$ and thus the ratio of compressibilities of kaolin to water $(1+b)$ equal to 0.016, 0.028, and 0.023. Apparently no great percentage accuracy is attainable for substances of such low compressibility as kaolin, since the much larger compressibility of the suspending liquid all but masks the slight effect of the compressibility of the particles themselves.

An alternative method of computation is available in case the concentration $\beta$ is not immediately obtainable. If in Eq. (1) we substitute

$$\beta=\frac{\rho_0-\rho_1}{\rho_2-\rho_1}$$

we may obtain this equation in the form

$$\frac{\rho(\rho-1)}{\gamma^2-\rho}=\frac{a}{b}.$$

The ratio $a/b$, denoting the ratio of the difference of density to the difference of compressibility, is a property of a suspension independent of the concentration of suspended substance, and can be found from density and velocity measurements on any one suspension and the suspending liquid.

This may be exemplified by data on some fresh horse blood obtained from the Naval Medical Research Institute. Blood can be considered to be a suspension of several types of blood corpuscles in a clear pale yellow fluid called the plasma. The red blood cells, or erythrocytes, with a diameter of about 8 microns, outnumber the other types of corpuscles by about 800 to 1, and consist of the red substance, hemoglobin, surrounded by a thin flexible membrane.

By allowing a quantity of fresh blood to settle, samples of clear plasma and concentrated red cells were obtained. Velocity and density measurements were made on these two end members, and also on samples of intermediate red cell concentration obtained by mutual dilution. This data is plotted in Fig. 5, with density, representing concentration of red cells, as abscissa. The lower portion of the plot shows the values of the ratio $a/b$ found from the velocity and density measurements. It is seen to have a constant value, within 3 percent, of $-0.433$. In order to obtain the compressibility of the corpuscles, from this value, we must know their density. This was found by centrifuging a sample of previously measured density. The percentage volume of corpuscular material (or hematocrit reading) yields the density of the corpuscles. This value, 1.091, together with the average value of $a/b$, and the compressibility of the plasma, $40.9\times10^{-12}$ cm$^2$ dyne$^{-1}$ (found from its density and velocity) give as a result the value of $34.1\times10^{-12}$ cm$^2$ dyne$^{-1}$ as the compressibility of the blood corpuscles. Thus, the corpuscles are found to be 16.7 percent less compressible than the plasma in which they are immersed.

## SUMMARY

A method is described by which the adiabatic compressibility of small particles can be found. It requires the measurement of the velocity of sound and of the density on both a suspension of the particles, and on the suspending liquid. The limitations of the method in regard to particle size are undetermined, but a coarse suspension such as blood seems to yield velocities (at a frequency of 1 megacycle) predictable by the simple theory outlined. The requirement of having a sample of the suspending liquid available for measurement is easily met by allowing the mixture to settle, or by centrifuging. It is necessary to control the temperature of the sample to 0.1°C, since the velocity change between suspension and the pure liquid is often slight. The accuracy obtainable is rather small for relatively incompressible solids, such as kaolin, but is probably fairly great for particles having a compressibility not too much smaller than that of the suspending liquid.

# End-Effect in Rotational Viscometers

CHARLES H. LINDSLEY AND EARL K. FISCHER
*Institute of Textile Technology, Charlottesville, Virginia*

(Received July 7, 1947)

In the classical equations for the traction on concentric cylinders by a viscous substance, it is assumed that traction on the top and bottom of the inner cylinder (bob) is negligible in comparison with that on the curved surface. In viscometers of practical dimensions, however, it is necessary to compensate for the end-traction or end-effect by adopting one of several expedients in design or by determining the magnitude of the effect and accounting for it in terms of increased length of bob.

In the experiments reported in this paper, the method of multiple bobs has been used, and the variation of end-effect with the following factors studied: (1) radius of bob; (2) clearance between cylindrical surfaces; (3) separation between bottom of bob and cup; and (4) viscosity.

The magnitude of the end-effect increases with radius and with clearance. For separations at the bottom greater than 1 cm and for viscosities above 1 poise the end-effect is nearly constant, but must be determined for each cup and bob combination. At lower viscosities the correction must be determined either by calibration with a standard liquid of about the same viscosity as the unknown, or by the multiple-bob method.

It is shown that trapping a layer of air beneath the bob is not effective in making the traction on the bottom negligible. The end-effect is almost as large for a bob with an open bottom as for a closed one.

When the method of multiple bobs is used, data with an accuracy of ±2 percent can be obtained without calibration. When the instrument constant is determined with standard liquids or computed from a value for the end-effect previously found, data of somewhat lower accuracy result, but the uncertainty should be within ±5 percent.

## INTRODUCTION

THE co-axial-cylinder viscometer is generally considered the most versatile instrument for the measurement of viscosity. It may be used on suspensions that exhibit anomalous flow properties as well as on Newtonian liquids, and it is suitable for determinations over a wide range of viscosities at variable rates of shear. The literature on instruments of this type is extensive, but relatively few studies have been made on several factors, inherent in the design, which seriously affect the validity of the measurement.

One of the most important of these factors is the contribution of the ends of the inner cylinder or bob to the total retarding force between the bob and the liquid surrounding it. This end-effect has not received the consideration it needs. Accordingly, the present investigation was undertaken with the following objectives: (1) critical evaluation of various methods for eliminating or measuring the end-effect; (2) experimental measurement of the end-effect for cups and bobs of several sizes; and (3) determination of the variation of the end-effect with viscosity and with clearance between cup and bob. To carry out these objectives a commercial Stormer viscometer was used after modification, but the results can be applied to any rotational viscometer equipped with cylindrical cups and bobs.

The mathematical analysis for the determination of viscosities in rational units by the method of rotating co-axial cylinders leads to the expression:[1]

$$T=4\pi\left[\frac{R_c^2\times R_b^2}{R_c^2-R_b^2}\right]h\eta\omega \tag{1}$$

where $T$ is the driving (or retarding) torque on the bob, $R_c$ the radius of the cup, $R_b$ the radius of the bob, $h$ the length of the bob, $\eta$ the viscosity of the liquid between cup and bob, and $\omega$ the speed of rotation at equilibrium. The radii and length being known, $T$ and $\omega$ are found experimentally, and hence $\eta$ may be calculated. For materials that exhibit plastic-flow properties the equation given by Reiner and Riwlin[2] is now considered to account fairly well for the results obtained.

---

[1] See, for example, H. Lamb, *Hydrodynamics* (Cambridge University Press, London, 1932), 6th edition, p. 587.

[2] M. Reiner and R. Riwlin, Kolloid Zeits. **43**, 1–5 (1927); M. Reiner, J. Rheology **1**, 5–10 (1929); H. Green, Ind. Eng. Chem. Anal. Ed. **14**, 576–585 (1942), has examined this equation critically and found experimental evidence to support it.

## METHODS FOR CORRECTING FOR TRACTION ON BOB ENDS

In the derivation of Eq. (1) it is assumed that the traction of the viscous liquid on the top and bottom of the inner cylinder is negligible in comparison with the traction on the curved surface. No adequate mathematical treatment has yet been given of the contribution of the ends to the total traction on the bob. It is necessary, therefore, to take it into account in one of the following ways: (1) use of a very long bob of small radius, with small clearance between cup and bob, to make the correction negligible; (2) calibration of the instrument with liquids of known viscosity, thus absorbing the end-effect in the instrument constant; (3) reduction of the relative motion of the liquid above and below the ends of the bob by mechanical expedients; and (4) measurement of the effect experimentally in order to include it in the calculations of viscosity and yield value.

The first method imposes many difficulties in the design of suitable apparatus. The second method is widely used, but one should be aware of the fact that the correction (and hence the instrument constant) changes with viscosity as will be shown in the experimental portion of this paper. Mooney and Ewart[3] devised an ingenious way of including the correction in the instrument constant: the bottom of their bob was a very flat cone (half angle 87.03°) with the apex resting on the center of the bottom of the cup. With this arrangement they showed that the rate of shear across the bottom was the same as that along the cylindrical surface; the constant obtained on calibration should therefore hold much better over a wide range of rotational speeds and viscosities than is the case with a bob of conventional type.

The third method was used in the instruments of Čouette,[4] Drew,[5] and Hatschek.[6] Stationary guard rings of the same radius as the inner (suspended) bob were placed as closely as possible above and below the bob. These rings prevented the liquid at the ends from rotating as the cup turned without interfering with the traction on the cylindrical surface, and thus eliminated the end-effect. Such instruments require machining and alignment of high precision and are difficult to clean. They are, therefore, suitable for standardization and research but are hardly practical for routine control use.

A widely used method, purporting to eliminate or greatly reduce the end-effect is to make the bottom of the bob slightly concave so that air is trapped in the cavity when the bob is immersed in the liquid to be measured (Fig. 1C). The assumption is that since the bottom of the bob rotates against a layer of air, of very low viscosity, the traction will be negligible. We have not found any experimental evidence in the literature that validates this assumption. Mercier[7] provided a raised portion in the bottom of the cup that almost closed the recess in the bob in order to exclude the liquid. His experimental data do now show whether he succeeded, however.[8] Nevertheless, Mooney and Ewart[3] refer to Mercier's "ingenious design (which) practically eliminates shearing stresses over the bottom surface." Goodeve[9] describes an instrument in which the cylindrical cup and bob are replaced by co-axial truncated cones, the ends of which are dished out. In the first paper he states that this reduces the end-effects to negligible values, and calculates viscosities using the measured dimensions of the cones; yet he gives no evidence to show that valid results are so obtained. In the second paper he only claims that the drag on the lower end is reduced by the trapped air, adding that "the reduction is not as large as might be expected because surface tension draws the fluid up into the rim"; he therefore calibrates the instrument with liquids of known viscosity,

---

[3] M. Mooney and R. H. Ewart, Physics 5, 350–354 (1934).

[4] M. M. Couette, Ann. de chimie 21, 433–510 (1890).

[5] E. R. Drew, Phys. Rev. (First Series) 12, 114–120 (1901).

[6] E. Hatschek, *Viscosity of Liquids* (Bell, London, 1928).

[7] P. Mercier, J. Rheology 3, 391–412 (1932).

[8] There is a strange gap in Mercier's discussion of his experimental results (reference 7, p. 405). It would appear that a considerable portion of the discussion had been omitted from the printed paper. The graphical presentation of his data gives no indication of success or failure in eliminating the end-effect. Furthermore, the graph indicates that the instrument was used at speeds as high as 4000 r.p.m. It is remarkable, to say the least, that he was able to operate this type of instrument, with a clearance of only 1 mm between cup and bob, at such speeds without getting turbulence and centrifugal effects that would cause deviations from a linear relationship.

[9] C. F. Goodeve and G. W. Whitfield, Trans. Faraday Soc. 34, 511–520 (1938), and C. F. Goodeve, J. Sci. Inst. 16, 19–27 (1939).

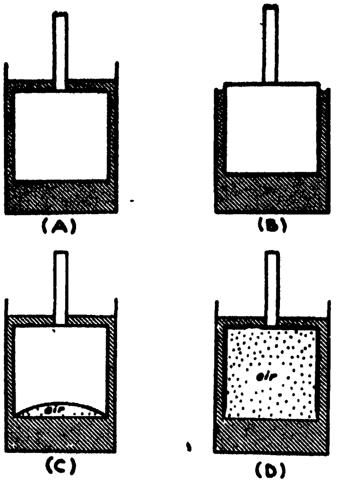


Fig. 1. Types of Cylindrical Bobs: A. Closed bob, completely immersed; B. Closed bob, immersed to different depths; C. Bob with concave bottom; D. Open bob.

including the end-effect in the empirically determined instrument constant.

In the absence of any sound experimental evidence of the effectiveness of this method in eliminating end-effect, we have studied the behavior of cylindrical bobs with open bottoms and find an effect almost as large as for a closed bob of the same dimensions. The effect as measured includes, of course, the drag on the top, which is the same for both bobs, but this can be shown to be smaller than the bottom effect. The method therefore does not accomplish the purpose for which it was designed and introduces a new uncertainty, namely, the extent to which the liquid does rise inside the bob.

The fourth method of evaluating the end-effect has been used by several investigators. In principle the traction on two bobs of the same radius but different lengths is measured, and by a difference calculation, the effect of the ends may be determined. Searle[10] designed an instrument with a long inner cylinder, the height of liquid about it being changed by raising or lowering the surrounding cup (cf. Fig. 1B). The time $T$ for one revolution was determined for several values of a falling weight $W$ used to turn the inner bob. For any given depth of immersion $h$, the product $WT$ was constant, and when $WT$ was plotted against $h$ for several depths, a straight line was obtained with a negative intercept $h_0$ on the $h$-axis. This intercept may be interpreted as an increment in bob length; that is, a bob having no end-effect, when immersed to a depth $h+h_0$, would exert the same traction as the actual bob immersed to depth $h$.

Lillie[11] made a careful study of the end-effect with an instrument designed primarily to measure viscosities of molten glass. He used two sets of five bobs each, those within each set having a constant radius but differing in length. Substitution in Eq. (1) for the dimensions of cup and bob and for the experimental quantities $T$ and $\omega$ gave a value for the "apparent viscosity" for each bob used with a given liquid. When these values were plotted against the reciprocals of the corresponding bob lengths ($h$), a straight line was obtained; its intercept on the axis $1/h=0$ (i.e. for a bob of infinite length) gave what Lillie termed the "true viscosity." Substitution of this value for $\eta$ in Eq. (1), $h$ being considered unknown, gave a value $h^*$ that might be called the effective length of each bob. The difference between this value and the actual length, $h^*-h=h_0$, was a measure of the end-effect quite analogous to $h_0$ as found by Searle.

## EXPERIMENTAL PROCEDURES

A commercial model Stormer viscometer was used after some important modifications had been made. The water jacket supplied with the

Table I. Dimensions of cups and bobs.

| | No. | Radius, cm | Height, cm |
|---|---|---|---|
| Cup | 1 | 1.746 (inside) | 5.2 (inside) |
| | 2 | 2.342 (inside) | 4.9 (inside) |
| | 4 | 2.181 (inside) | 5.0 (inside) |
| | 5 | 2.539 (inside) | 5.3 (inside) |
| | 6 | 2.800 (inside) | 5.3 (inside) |
| Bob | 1 | 1.565 | 3.53 |
| | 5 | 1.566 | 2.00 |
| | 6 | 1.565 | 1.00 |
| | 7 | 1.566 | 0.08 |
| | 14 (open) | 1.590 | 3.49 |
| | 3 | 1.269 | 3.51 |
| Bob | 4 | 2.104 | 3.49 |
| | 10 | 2.100 | 1.99 |
| | 11 | 2.100 | 1.00 |
| | 12 | 2.100 | 0.08 |
| | 15 (open) | 2.103 | 3.46 |

---

[10] G. F. C. Searle, Proc. Camb. Phil. Soc. 16, 600–606 (1912). For an earlier application of the method, see L. E. Gurney, Phys. Rev. (First Series) 26, 98–120 (1908).

[11] H. R. Lillie, J. Am. Cer. Soc. 12, 505–515 (1929); Phys. Rev. 36, 347–362 (1930).

instrument was wound with resistance wire (18 ft. of 22-gauge Nichrome) and then with several layers of asbestos paper. Current for this heating coil was supplied through a Variac and controlled by a fixed-contact mercury thermoregulator and a suitable relay. At 30°C the temperature could be held to within ±0.1°. Other thermoregulators were available to give temperatures up to 90°C. At high temperatures, however, enough heat may be lost through the shaft of the bob to lower the temperature of the bob and of the liquid in contact with it and thus introduce appreciable error. A plastic bob with brass shaft was made for use at high temperatures; the heat transfer at the surface of such a bob is far less than with an all-metal bob.

The construction of the cup and bob supplied with the instrument makes them unsuited for determination of viscosity in absolute units. Several cylindrical cups and bobs were machined from brass of such sizes that a variety of rates of shear for any given rate of rotation was possible. The dimensions are shown in Table I. The larger bobs were hollow to reduce inertia and weight on the bearing. Duplicates of two bobs were made with the lower ends left open to permit a study of the end-effect for concave bobs (Fig. 1D).

In order to evaluate the results obtained by the several procedures followed, four viscosity samples, certified at 30°C, were obtained from the National Bureau of Standards, the viscosity ranging from 0.2648 to 153.9 poises. A description of these standards has been given recently by Swindells.[12] In addition, a series of four bodied linseed oils (litho-varnishes) were obtained to give better coverage of the range.

TABLE II. Calculation of viscosity and end-effect for sample N-13 at 30°C. (Bobs 1, 5, 6, 7 in cup 1.) $K/r^2 = 99.0/12.46 = 7.95$.

| | Bob 1 | Bob 5 | Bob 6 | Bob 7 |
|---|---|---|---|---|
| Bob length, $h$ | 3.53 | 2.00 | 1.00 | 0.08 |
| Load/r.p.m., $P$ | 3.18 | 1.00 | 1.17 | 0.42 |
| $\Delta P$ | — | 1.18 | 2.01 | 2.76 |
| $\Delta h$ | — | 1.53 | 2.53 | 3.45 |
| $\eta$, poises | — | 6.14 | 6.31 | 6.35 |
| Average value of $\eta$ | | 6.27 | | |
| Effective length, $h^*$ (cm) | 4.13 | 2.60 | 1.52 | 0.56 |
| End-effect, $h_0 = h^* - h$ (cm) | 0.60 | 0.60 | 0.52 | 0.48 |

TABLE III. Experimentally determined viscosities of certified samples at 30°C.

| Sample | Experimental viscosity (poises) | | | | Certified viscosity | Deviation (%) |
|---|---|---|---|---|---|---|
| | Bob combination | | | Av. of first two | | |
| | 1-5 | 1-6 | 1-7 | | | |
| K-5 | 0.255 | 0.276 | 0.261 | 0.265 | 0.2648 | 0.0 |
| M-11 | 1.12 | 1.17 | 1.19 | 1.15 | 1.166 | −1.4 |
| N-13 | 6.14 | 6.31 | 6.35 | 6.22 | 6.124 | 1.5 |
| OB-1 | 156. | 158. | 154. | 157. | 153.9 | 2.0 |

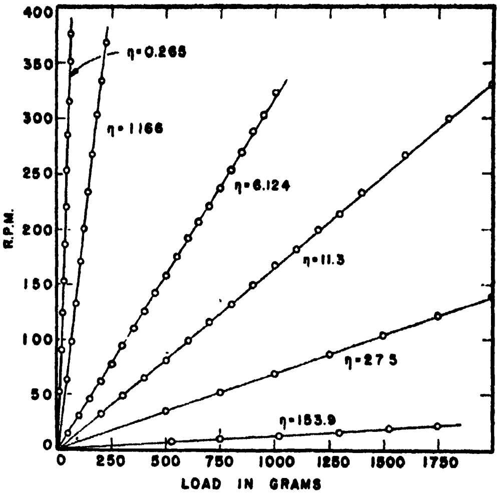


FIG. 2. Experimental values for load and r.p.m. for oils showing Newtonian flow.

The procedure in using the viscometer follows. The cup and bob were carefully aligned, the bob being raised about 1.3 cm above the bottom of the cup. The cup was then filled to the top with the sample to be measured. After the sample had reached 30°C (the temperature at which all measurements here reported were made), at least ten measurements of the time of rotation with different loads were taken. The ratio of load to r.p.m. was calculated and found to be nearly constant for each sample so long as the rotation was held below 400 r.p.m. (Fig. 2). This constancy indicated that the oils were Newtonian in their flow properties.

## THE SINGLE-BOB METHOD

Some preliminary measurements were made with one of the standard samples (N-13) following the method of Searle, that is, by immersion of one bob to varying depths. The intercept was found to be 0.29 cm with bob 1 and cup 1. With this value for the end-effect added to the actual

---

[12] J. F. Swindells, J. Colloid Sci. 2, 177–84 (1947).

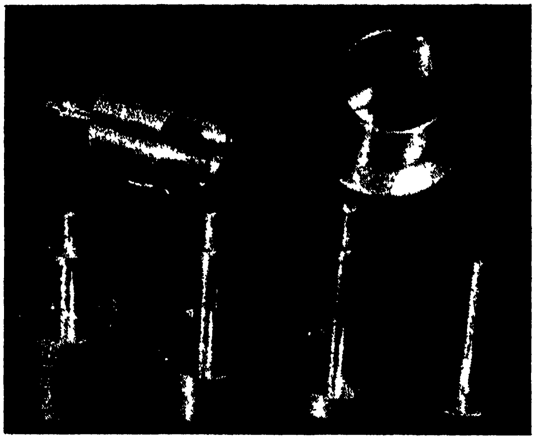


FIG. 3. Four bobs of equal radius but different height; open bob; cup.

length of the bob, the viscosity of the sample was found by Eq. (1) to be 6.36 poises, in approximate agreement with the certified value of 6.12 poises. The principal source of error in the method lies in the inaccuracy with which depth of immersion of the bob is measured. In spite of this drawback, the method affords a rapid and easy means of obtaining viscosities in cgs units with an accuracy of about 5 percent, and only one machined cup and bob are needed.

## THE MULTIPLE-BOB METHOD

For more accurate results a method similar to that used by Lillie is preferred: the end-effect may be first eliminated and afterwards evaluated (if desired) by using a set of bobs equal in radius but different in length (Fig. 3). For each bob the ratio of load to r.p.m. is determined in the usual way. From the ratios for any pair the viscosity is calculated by the following modification of Eq. (1). Let $L$ be the torque exerted by the liquid against the ends of the bob, and $T$ the total torque exerted against the ends and cylindrical surface. Equation (1) may then be written for two bobs of length $h_1$ and $h_2$:

$$T_1=4\pi r^2 h_1\eta\omega_1+L$$

and

$$T_2=4\pi r^2 h_2\eta\omega_2+L$$

where

$$r^2=R_c^2\times R_b^2/(R_c^2-R_b^2).$$

If $T_1$ and $T_2$ are chosen so that $\omega_1=\omega_2$, these equations may be solved for $\eta$:

$$\eta=\frac{T_1/\omega_1-T_2/\omega_2}{4\pi r^2(h_1-h_2)}. \tag{2}$$

Since $T/\omega$ is constant for all values of $\omega$ for a Newtonian liquid, the average values of $T_1/\omega_1$ and $T_2/\omega_2$ may be used, and not just those for which $\omega_1=\omega_2$. With the Stormer viscometer the weight or load may be substituted for torque thus:

$$T=Wgl$$

where $g$ is the acceleration of gravity and $l$ the effective length of the arm on which the weight acts. It is also more convenient to express angular velocity in r.p.m. than in radians per second:

$$\omega=\frac{2\pi}{60}\cdot\text{r.p.m.}$$

If we designate by $P$ the experimentally determined ratio $W$/r.p.m., Eq. (2) becomes·

$$\eta=\frac{K}{r^2}\times\frac{\Delta P}{\Delta h} \tag{3}$$

where

$$K=\frac{60gl}{8\pi^2}.$$

For the Stormer viscometer $l$ is the radius of the drum on which the cord is wound plus the cord radius, divided by the gear ratio. In our instrument $l=1.456/11$, and the constant $K$ has the value 99.0.

To illustrate the use of this equation, data obtained with bobs 1, 5, 6, and 7 in cup 1 are shown in Table II. For $\Delta P$ and $\Delta h$ the differences between the $P$ and $h$ values for bob 1 and each of the other bobs are taken. (Three other combinations are possible but are not so useful.) The average of the three calculated viscosities is 6.27 poises, in fair agreement with the certified value of 6.124. Since the relative effect of the ends is less for bobs 1 and 5 than for bobs 6 and 7, the first two should yield a more accurate value for $\eta$. This expectation is borne out by the agreement with the certified value.

In Table III are given the viscosities of four of the certified samples determined by the method just illustrated in detail. Although the viscosities calculated using bob 7 were not included in the averages, to have done so would not have changed the results significantly. The figures indicate that viscosities over a wide range may be determined in absolute units with an accuracy of about 2 percent.

With the value of $\eta$ so determined the same experimental values of $P$ are used to calculate the effective length of the bob $h^*$, and from it the end-effect $h_0$ expressed as increment in length, by means of the equation:

$$h^* = \frac{K}{r^2} \times \frac{P}{\eta} = h + h_0. \tag{4}$$

Such calculations are illustrated in the last two lines of Table II. It is interesting that even for a very thin bob (No. 7) for which the traction on the end is nearly 90 percent of the total, the end-effect is nearly the same as for a bob more than 40 times longer.

One would expect the end-effect to increase as the bottom of the bob is brought closer to that of the cup; moving in the other direction, one would expect that above a certain point there would be no significant change with increasing separation. This behavior was demonstrated by determining the end-effect with a short bob (No. 5) at varying depths of immersion. The results are shown graphically in Fig. 4. Above 1 cm the variation of end-effect with depth is insignificant. The height of liquid above the top of the bob is not critical, so long as it is at least 2 mm.

A different approach to the measurement of end-effect was attempted by use of bobs open at the bottom as noted above. At the same time a check on the effect of the top of the bob was made by carrying out runs with the top just level with the top of the cup filled with liquid and with it submerged 5 mm below the surface of the liquid. The results are shown in Table IV. The second column shows the combined effect of top and bottom; the third column the effect of the bottom, while the difference (fourth column) shows the effect of the top. The uncertainty in the values for end-effect in this series of measurements is estimated as about 1 mm. The apparent inversion for the data of bob 1, cup 1 may be ascribed to this factor.

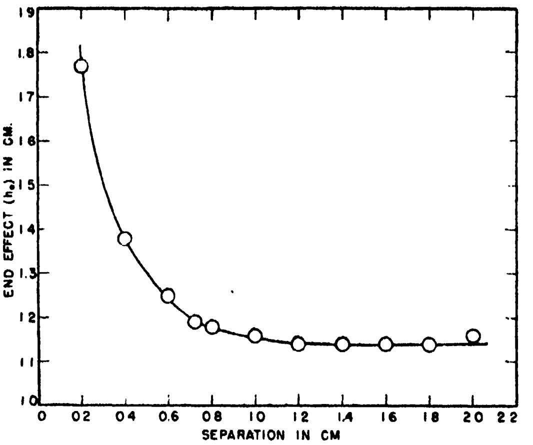


FIG. 4. Variation in end-effect with separation between bottom of bob and cup.

With the limitations on precision in such measurements in mind, the following observations may be made: (1) virtually the same traction is exerted at the bottom of the bob whether it is open or closed; (2) in general the bottom exerts a greater traction than the top, due probably to the greater freedom of the liquid above the bob to rotate with it. The first result was quite unexpected, and is contrary to much that is stated or implied in the literature and in the design of several viscometers. It was therefore checked by repeating the comparison of closed and open bobs in liquids differing markedly in viscosity. The results are shown in Table V. Since a number of variables are involved, the most valid comparisons are made for each pair of figures for open and closed bobs. In the last column in which end-effect is expressed as percent of effective bob length, it will be seen that although the results with the open bob are less

TABLE IV. End-effect for bobs closed and open at the bottom, with tops immersed or exposed.

| Cup and bob combination | | End-effect, cm Top immersed 0.5 cm | Top exposed | Difference |
|---|---|---|---|---|
| Cup 1 | Bob 1—closed | 0.6 | 0.4 | 0.2 |
| | Bob 14—open | 0.7 | 0.5 | 0.2 |
| Cup 2 | Bob 1—closed | 1.1 | 0.7 | 0.4 |
| | Bob 14—open | 1.0 | 0.5 | 0.5 |
| Cup 2 | Bob 4—closed | 0.7 | 0.3 | 0.4 |
| | Bob 15—open | 0.6 | 0.3 | 0.3 |

TABLE V. Comparison of end-effects for closed and open bobs.

| Cup no. | Clearance (cm) | Approx. viscosity (poises) | Effective bob length, $h^*$ Bob 1 (closed) $h=3.53$ cm | Effective bob length, $h^*$ Bob 14 (open) $h=3.48$ cm | End-effect, $h_0$ Bob 1 (closed) cm | End-effect, $h_0$ Bob 14 (open) cm | $100h_0/h^*$ Bob 1 (closed) (%) | $100h_0/h^*$ Bob 14 (open) (%) |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.18 | 1.17 | 4.10 | 4.04 | 0.57 | 0.56 | 16 | 16 |
| | | 6.12 | 4.07 | 3.95 | 0.54 | 0.47 | 15 | 14 |
| 4 | 0.60 | 1.17 | 4.46 | 4.17 | 0.93 | 0.69 | 26 | 20 |
| | | 6.12 | 4.25 | 4.06 | 0.72 | 0.58 | 20 | 17 |
| | | 27.5 | 4.48 | 4.18 | 0.95 | 0.70 | 27 | 20 |
| | | 154. | 3.97 | 3.67 | 0.44 | 0.19 | 13 | 5 |
| | | | Bob 4 (closed) $h=3.49$ cm | Bob 15 (open) $h=3.46$ cm | Bob 4 (closed) cm | Bob 15 (open) cm | Bob 4 (closed) (%) | Bob 15 (open) (%) |
| 2 | 0.24 | 1.17 | 4.04 | 3.91 | 0.55 | 0.45 | 16 | 13 |
| | | 6.12 | 3.92 | 3.81 | 0.43 | 0.35 | 12 | 10 |
| | | 27.5 | 4.31 | 4.20 | 0.80 | 0.74 | 23 | 21 |
| | | 154. | 4.12 | 4.04 | 0.63 | 0.58 | 18 | 17 |
| 3 | 0.17 | 1.17 | 4.00 | 3.84 | 0.51 | 0.38 | 15 | 11 |

than those with the closed bob, the reduction is not nearly so high as one would expect.

Since the end-effect as normally measured varied with separation between bottom of bob and cup, it was of interest to see if a similar change would be observed with an open bob. The results obtained are shown in Fig. 5: the end-effect is slightly less when the bob has no bottom, but otherwise is quite similar.

## VARIATION OF END-EFFECT WITH CLEARANCE AND VISCOSITY

The influence on the end-effect of the following factors was studied: radius of the bob, clearance between bob and cup, and viscosity of the liquid being measured. Table VI summarizes many determinations with different cup and bob combinations; most of the values given are averages of several measurements made with bobs of the same radius but different lengths (as illustrated by the data for $h_0$ in Table II) and with oils of viscosities between 1 and 154 poises. The precision of the measurements as shown by variation in $h_0$ for different oils appears low, amounting to as much as 25–30 percent, but this is because errors are magnified since $h_0$ is obtained by difference. For example, if the effective length of a bob 3.5 cm long is found to be 4.5 cm, the end-effect is 1.0 cm; then an error of 5 percent in $h^*$ makes an error of 22 percent in $h_0$. It is clear, however, that the effect increases with clearance and that, for a given clearance, the effect may be somewhat larger the larger the radius of the bob. In explanation, perhaps more disturbance is created in the liquid by the end of the bob with relatively large clearances than when these surfaces are close together. The same explanation may account for the lower precision attained at the larger clearances.

TABLE VI. Variation of end-effect with clearance between cup and bob for several bob radii.

| Bob radius, cm | Clearance, cm | End-effect, cm |
|---|---|---|
| 1.27 | 0.48 | 0.8 |
| | 1.08 | 1.2 |
| 1.57 | 0.18 | 0.6 |
| | 0.78 | 1.0 |
| | 0.97 | 1.4 |
| | 1.24 | 1.4 |
| 2.10 | 0.08 | 0.6 |
| | 0.24 | 0.8 |
| | 0.44 | 1.2 |
| | 0.70 | 1.4 |

TABLE VII. Variation of end-effect with viscosity (clearance 0.18 cm).

| Viscosity at 30°C | End-effect |
|---|---|
| 0.017 poise | 11.5 cm |
| 0.058 | 2.94 |
| 0.145 | 1.94 |
| 0.265 | 1.28 |
| 1.166 | 0.60 |
| 1.48 | 0.65 |
| 6.124 | 0.52 |
| 9.2 | 0.60 |
| 11.3 | 0.59 |
| 27.5 | 0.56 |
| 153.9 | 0.49 |

Table VII shows the variation of the end-effect with viscosity observed with the combination of cup 1 and bob 1. Entirely similar data were obtained with other combinations. For viscosities above 1 poise the effect is nearly constant (about 0.6 cm in the example shown). As one goes to lower viscosities, however, the end-effect becomes progressively greater. The greater the clearance between cup and bob, the more rapidly the end-effect increases as viscosity decreases. These facts too can be explained by assuming that at low viscosities the rotating ends of the bob interfere with laminar motion of liquid between the cylindrical surfaces.

Even with these variations, if only those combinations of cup and bob are used for which the clearance is small (say, less than 0.8 cm), a value for the end-effect may be found which permits determination of viscosity in poises over a wide range. This is illustrated by the data in Table VIII. Viscosities were calculated from Eq. (4), with the best value for $h_0$ and hence for $h^*$ determined by the multiple-bob method described above. It should be noted that at no point in the determination of the viscosities in Table VIII have the certified or known values for the samples been used. Over this range the correction for end-effect may be applied to give viscosities with an accuracy within ±5 percent.

## DISCUSSION

The advantages of the co-axial cylinder viscometer are so numerous that it deserves wider use for industrial measurements. As it is commonly used, however, the data obtained cannot be expressed in rational units and, as a consequence, the technical literature is cluttered with meaningless measurements. The unsatisfactory state of instrumentation in this field is in part responsible, but the mechanical changes necessary to convert a standard commercial Stormer viscometer are comparatively simple and can be made at relatively small cost. The modified instrument makes possible the measurement of viscosities based on accepted physical principles and yields data which are dimensionally consistent and experimentally reproducible. The precautions necessary for satisfactory operation —aside from such obvious matters as temperature control, centering of bob and cup, etc.— are to a large extent elucidated by the experimental data presented above.

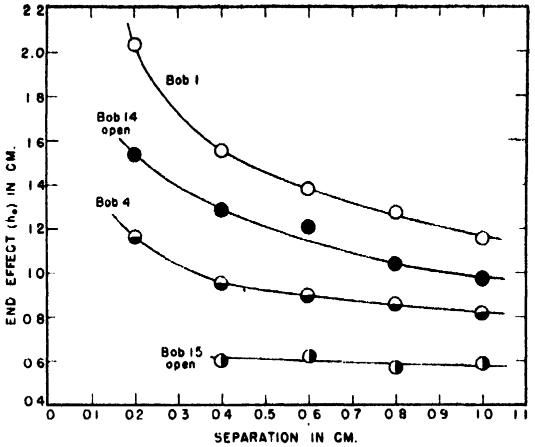


FIG. 5. Variation in end-effect for closed and open bobs.

In making cups and bobs for the Stormer viscometer, the important dimensions are the

TABLE VIII. Viscosities determined by correcting bob length for end-effect. $\eta = (K/r^2h^*) \times (\text{Load/r.p.m.})$.

| Combination | | Cup 1—bob 1 | | Cup 2—bob 1 | | Cup 2—bob 4 | |
|---|---|---|---|---|---|---|---|
| $K/r^2$ | | 7.95 | | 22.4 | | 4.32 | |
| Clearance | | 0.18 | | 0.78 | | 0.24 | |
| Correction, $h_0$ | | 0.58 | | 1.00 | | 0.82 | |
| Effective length, $h^*$ | | 4.11 | | 4.53 | | 4.31 | |
| Sample | Viscosity, known | Viscosity, detmd. | Deviation (%) | Viscosity, detmd. | Deviation (%) | Viscosity, detmd. | Deviation (%) |
| NBS: M-11 | 1.166 | 1.16 | −0.5 | — | — | 1.15 | −1.4 |
| N-13 | 6.124 | 5.90 | −3.7 | 5.97 | −2.5 | 6.00 | −2.0 |
| OB-1 | 153.9 | 151. | −1.9 | — | — | 148. | −3.8 |
| Litho-varnishes | | | | | | | |
| No. 00000 | 1.5 | 1.49 | 0 | 1.52 | 0. | 1.45 | 0. |
| No. 1Trans. | 9.2 | 9.3 | +1.1 | 9.0 | −2.2 | 9.3 | +1.1 |
| No. 1 Reg. | 11.6 | 11.7 | +0.9 | 12.0 | +3.4 | 11.6 | 0.0 |
| No. 3 Reg. | 27.5 | 27.9 | +1.5 | 27.0 | −1.8 | 27.5 | 0.0 |

inside diameter of the cup and the diameter and length of the bob. Machining tolerances should be within 0.002 cm (0.001 in.). For most measurements two cups and bobs, making available three clearances, will be found enough; we have found cups 1 and 2 and bobs 1 and 4 most useful.

Because of the nature of the study we were making, the end-effects were determined without any reference to standard viscosity samples except for final confirmation. For this purpose the method of multiple bobs is recommended. For most applications and industrial uses, however, a simple calibration with standard samples is more convenient and entirely satisfactory. It is enough to determine $P$, the ratio of load to r.p.m., as described above, and substitute for it and the known value of $\eta$ in the equation $\eta = K_v \cdot P$ in order to determine the constant $K_v$. This constant combines the end-effect with all the various instrument constants, dimensions of cup and bob, etc.

Within the range from 1 to 150 poises (and probably considerably higher) calibration with a single standard sample will be found sufficiently accurate for most purposes. Below 1 poise, however, since the end-effect has been shown to become increasingly large, calibration should be made with a standard the viscosity of which is close to that of the material to be measured. If this is not possible, the multiple-bob method should be used.

A separation of at least 1 cm should be kept between the bottom of the bob and cup, and an overlay of about 5 mm of liquid above the top of the bob is desirable. In addition to constancy in end-effect, one advantage of the bob immersed well below the liquid level when measuring solutions or suspensions is that evaporation of all except the most volatile solvents introduces small or negligible effects during the course of measurement.

## ACKNOWLEDGMENT

The authors are indebted to Mrs. Dorothy P. Stubbs and Mr. Samuel E. Miller who made many of the experimental measurements.

---

# The Theory of Disk-Loaded Wave Guides

E. L. Chu* and W. W. Hansen
*Department of Physics, Stanford University, Stanford University, California*

(Received July 11, 1947)

The properties of circular wave guides loaded with apertured disks are discussed both qualitatively and quantitatively. Formulae and curves are given for various quantities including the wave and group velocities, the attenuation, and the power flow.

## INTRODUCTION

IN the normal circular wave guide shown in Fig. 1a, the running waves of field propagate with a phase velocity greater than that of light. By the introduction of apertured disks, as shown in Fig. 1b, which also defines some of our notation, this velocity can be reduced as much as desired. In addition to altering the phase velocity, the loading disks change the group velocity and, two parameters being available, the two velocities can be controlled independently. Finally, the periodic spacing of the disks introduces a high frequency cut-off, in addition to the low frequency one already present, so that the loaded structure has many of the characteristics of a band pass filter. We propose to report here certain calculations on the properties of such structures, which we call disk-loaded wave guides.

While the main formulae we use are of general interest, some special formulae and many of the numerical examples are developed to fit the special circumstances of an important possible use of disk-loaded guides—namely, that of accelerating electrons. Indeed, interest in this problem has been the primary motive for developing the theory that follows.

Before describing the details of the calcula-

---

[*] On leave from the Institute of Physics, Academia Sinica, Shanghai, China.

tions, it will be useful to consider several qualitative descriptions of the operation of such a structure.

One point of view regards the disk-loaded guide as derived from an infinite sequence of end-to-end cylindrical cavity resonators by opening up "coupling holes" between adjacent cavities. Then if we consider the behavior at various frequencies, as is done in Fig. 2, we start at zero frequency with Fig. 2a showing the electric-field lines terminating on the disks and decreasing rapidly in strength as we go from left to right. Here we are below the low frequency cut-off, the device acts as an attenuator, and curvature of electric-field functions toward the axis, when going in the $r$ direction, is canceled by curvature away from the axis in the $z$ direction to give the zero total curvature required at zero frequency. As we increase the frequency, the curvature away from the $z$ axis decreases until at the resonant frequency of the uncoupled resonators without holes ($\lambda = 2.61b$) the fields are in phase everywhere in the guide, there is no attenuation, and the field curvature is all in the $r$ direction, as shown in Fig. 2b. This field configuration may be regarded as obtained by removing disks of radius $a$ from the inter-resonator boundaries of a sequence of resonators all oscillating in phase, such removal not altering the fields since the currents on opposite sides of any removed disk are equal and opposite. As we further increase the frequency, so running the cavities off resonance, it becomes necessary to supply reactive power to each cavity. This comes through the coupling hole from the adjacent cavity which must then be of somewhat different phase. The resultant field pattern is sketched in Fig. 2c and it must now be remembered that this represents a running wave. As the frequency is further increased, the phase shift per section increases until it becomes 180° at which point the fields are as shown in Fig. 2d. As the frequency is further increased, more phase shift is impossible, attenuation again sets in and the fields are as in Fig. 2e. Thus we see that the two cut-off frequencies correspond to the two modes of cavity oscillation shown in Figs. 2f and 2g.

A second point of view regards the disk-loaded guide as derived from an ordinary guide by the introduction of loading disks which may at first be taken as having small radial extent. From this point of view the fields of Fig. 2a are obvious and Fig. 2b is nearly so; we merely have to remark that, at cut-off, the disks do not influence the electric field since they are at right angles thereto. As we enter the pass region, the disks do not suppress the propagation so long as the disks are closer together than half a loaded-guide wave-length, for in such a case the reflections from succeeding disks cannot add up but must, over a long distance, cancel. This ceases to be so when the guide wave-length decreases to the point that the disks are a half-wave apart, for then the reflections add and transmission is impossible. We then enter the attenuation region as in Fig. 2d.

The third point of view, and the one most closely connected with the mathematics here used to reduce qualitative ideas to quantitative, starts with the fields, and more particularly the fields near the axis, as given, and seeks a structure that fits the fields, as opposed to the above, and usual, procedure of starting with a structure and finding the fields.

Thus we start by requiring a field which, for $r<a$, is finite, axially symmetric, and represents a wave running in the $z$ direction with (real) wave number $k_3$. These conditions completely specify the wave functions—the only possible solution of Maxwell's equations is

$$\begin{aligned} E_z &= E_0 J_0(k_1 r) e^{-jk_3 z}, \\ E_r &= E_0 j(k_3/k_1) J_1(k_1 r) e^{-jk_3 z}, \\ B_\phi &= E_0 j(k/k_1) J_1(k_1 r) e^{-jk_3 z}, \end{aligned} \tag{1}$$

with all other components zero. Here $E_0$ is a constant, we use Gaussian units and a positive time factor, and

$$k_1^2 + k_3^2 = k^2 = 4\pi^2/\lambda^2.$$

It would appear that since $k_3$, which regulates the wave velocity, is unrestricted we can make waves which go with any velocity whatever. But we note that, the velocity being given by $v/c = k/k_3$, a velocity less than that of light implies a $k_3$ which is greater than $k$, and so an imaginary value of $k_1$. While the above are still solutions, their characteristics change when $k_1$ becomes imaginary, and we must investigate their properties for both real and imaginary $k_1$.

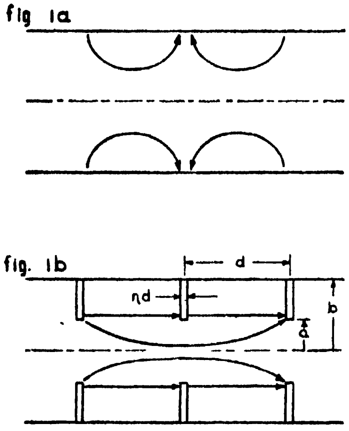


FIG. 1. The sketch of Fig. 1a shows the configuration of electric-field lines in an unloaded circular wave guide, while Fig. 1b shows the fields in a guide loaded by means of apertured disks, and also defines the notation used in describing the guide dimensions. The special case shown has the loads one-quarter wave apart—any other spacing is, of course, possible.

For real $k_1$ the Bessel function is oscillatory so that $E_z$ passes through zero for an infinite number of values of $r$. A qualitative field plot for this case is given in Fig. 3a. When $k_1$ is positive imaginary, however, $J_0$ and $J_1$ are monotonic increasing functions of $r$, so that the $E$ lines cannot reverse directions with the result that the field is qualitatively as shown in Fig. 3b.

Now, having specified the fields, we ask what metallic boundaries we can introduce to confine them to a finite region in space. In the case of fields like those of Fig. 3a one answer is well known: we can introduce a conducting tube in the position indicated by the dotted lines, in which case the electric field is everywhere perpendicular to the conductor and the currents carried by the conductor support the discontinuity in magnetic field. But this is impossible for the fields of Fig. 3b, for they never become perpendicular to the axis.

But another, and more general, possibility becomes apparent if we consider a special type of boundary. If we use a conducting sheet, the correct boundary conditions are $E_{\tan}=0$, $B_{\tan}\neq 0$, or $E/B=0$. But if we now add to the sheet a number of closely spaced conducting fins, each $\lambda/4$ long and perpendicular to the main sheet, we have a boundary which will sustain an electric field without a magnetic field since each pair of adjacent fins acts as a quarter-wave resonant line. The correct boundary conditions are then $E/B=\infty$. Moreover, it is easy to see that by choosing a fin depth between zero and $\lambda/2$, it is possible to obtain any desired value of $E/B$.

Thus, by using such a finned boundary we can contain any field and in particular the one assumed above. Thus we are led to the disk-loaded guide, and conclude that waves in such a structure can be made to run with any velocity desired.

In carrying out an analysis along the above lines it is plain that the results will approach exactness as the fins get closer and the phase shift per fin approaches zero. On the other hand, as the fin spacing is increased no qualitative changes occur until the phase change per fin becomes $\pi$, at which point the approach breaks down completely. This corresponds, of course, to the high frequency cut-off.

We can now conclude our introductory remarks by specifying the quantities given and to be computed and the approximations to be made.

The six quantities $a$, $b$, $d$, $\eta d$, the skin depth $\delta$ characterizing the guide material, and $\lambda$, may be regarded as given, and may obviously be reduced to five by measuring lengths in terms of $\lambda$. These five we take as $ka$, $kb$, $kd$, $\eta$, and $\lambda/\delta$. Usually $\lambda/\delta$ is large, in which case it may be

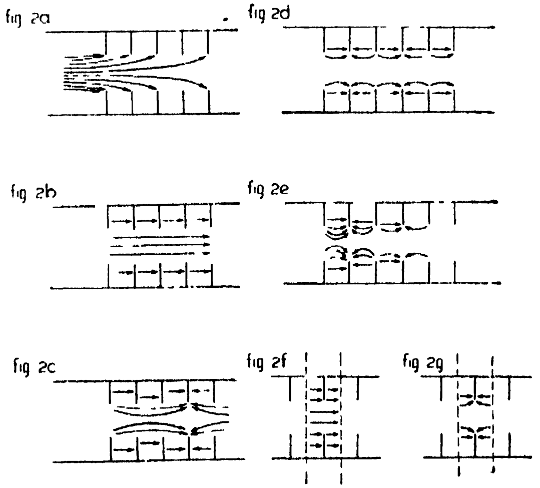


FIG. 2. Sketches a–e show the qualitative behavior of the $E$ lines in a loaded guide as the frequency is raised from zero to above the high frequency cut-off. Illustrations f and g show how a cavity can be constructed from a section of the guide by the addition of conducting planes, shown as dotted lines. The two indicated modes of oscillation of this cavity correspond to the low and high frequency edges of the pass band for the loaded guide.

disregarded in wave-length calculations and will appear only in loss calculations. Moreover, we will often assume that $kd$ is negligibly small, in which case there remain three quantities $ka$, $kb$, and $\eta$. Further, $\eta$ may sometimes be negligible also.

In principle there are only two quantities to compute, the wave number in the $z$ direction, which we call $k_3$ and which would be called the propagation constant in filter theory, and a characteristic impedance $Z_i$, to be defined more precisely later. As to $k_3$, a slightly more convenient quantity is $k_3/k = 1/\beta$, where the wave velocity is $\beta_c$. We also note that $k_3$ is complex so that we must compute both real and imaginary parts, the latter being a measure of the attenuation.

Although $k_3$ and $Z_i$ contain all the information, it is useful to compute also certain related quantities. These will be defined as the occasion arises but we may mention that the group velocity and the axial field for a given power are important, and that it is useful to compute not only the attenuation but the $Q$.

As to the approximations involved, they depend on the calculational approach used. As indicated in the qualitative discussion, at least three procedures are possible. First, we can treat the structure as a series of resonators perturbed by the coupling through the center holes. For such a calculation we must have $a \ll b$, $a \ll d$. Second, we may consider the loading to be nearly continuous, in which case we must have $d \ll a$, $d \ll b-a$. Third, we may consider a cylindrical wave guide, perturbed by the addition of loading disks. For this calculation we take $b-a \ll d$, $b-a \ll b$. For the computation of $k_3$ we have used all three methods. For the attenuation, impedance, and other quantities which are less critical we have usually used approximations based on the second approach. Thus it happens that the numerical majority of our results depend on this second approach, which will now be explained in detail.

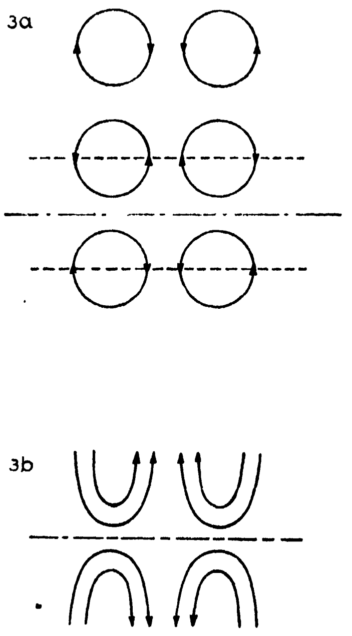


fig. 3a

fig. 3b

FIG. 3. Sketch a shows qualitatively the shape of $E$ lines given by Eq. (1) for real values of $k_1$, while b shows them for imaginary $k_1$. In the former case a cylindrical conducting boundary can be introduced at the radius indicated by the dashed lines.

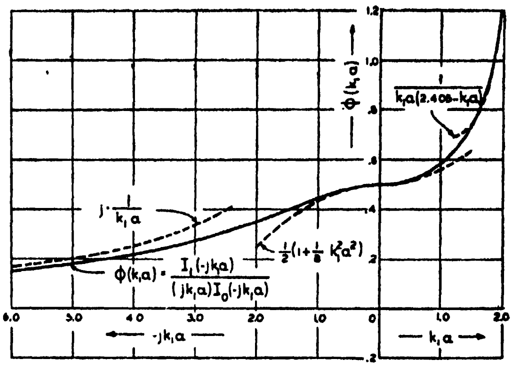


FIG. 4. This shows $\phi$ as a function of $k_1a$, real values of the argument being on the right-hand side, and imaginary values on the left. Various approximate formulae are shown in dotted lines.

## CALCULATIONS

### 1. Phase and Group Velocity

In (1) we have already given the fields to be used for $r < a$. For $a < r < b$ we take

$$E_z = CE_0Z_0(kr)e^{-ik_3(z)},$$

$$E_r = 0, \tag{2}$$

$$B_\phi = CE_0jZ_1(kr)e^{-jk_3(z)},$$

with $C$ a constant and $Z_0$ a Bessel function that is zero at $r=b$,

$$Z_0 = J_0(kr)N_0(kb) - N_0(kr)J_0(kb), \tag{3}$$

and

$$Z_1 = J_1(kr)N_0(kb) - N_1(kr)J_0(kb). \qquad (4)$$

Here $(z)$ is equal[1] to zero when $-d/2<z<d/2$, to $d$ for $d/2<z<3d/2$, etc., with the loading disks at $z=d/2$, $3d/2\cdots$. Otherwise said, $(z)$ is equal to the value of $z$ midway between the two nearest loading disks. Also, the fields are to be zero in the space occupied by the disks.

Plainly these are solutions of Maxwell's equations, and they satisfy the boundary conditions on the cylinder $r=b$ and on the plane surfaces of the disks. To join this function to that for $r<a$, we (a) make the values of $B_\phi$ match at $r=a$, and (b) make $(1-\eta)E_z$ from (2) at $r=a$ equal $E_z$ from (1). To do this we must adjust the constant $C$ and also find a $k_1$ such that the equation

$$\phi(k_1a)=\frac{1}{k_1a}\frac{J_1(k_1a)}{J_0(k_1a)}=\frac{1}{(1-\eta)ka}\frac{J_1(ka)N_0(kb)-N_1(ka)J_0(kb)}{J_0(ka)N_0(kb)-N_0(ka)J_0(kb)}=\frac{1}{1-\eta}\alpha(kd,\,kb) \qquad (5)$$

is satisfied.

This makes $B_\phi$ continuous at $r=a$, $z=(z)$, and the average value of $E_z$ likewise, and this is the best that can be done with these approximate functions. Besides the fact that $E_z$ and $B_\phi$ are not continuous at $r=a$, for general values of $z$, we note $E_r$ is finite for $r<a$ and zero for $a<r$. This is also correct on the average, for charge accumulates on the edges of the loading disks because of the difference in current from section to section.

The above is easy to understand, and is certainly right in the limit $k_3d\rightarrow 0$. If $k_3d$ is finite, the correct procedure would be to take $E_z$ at $r=a$ to be of the form $\cos k_3zf[z-(z)]$, where $f$ is some suitable function of the distance from the

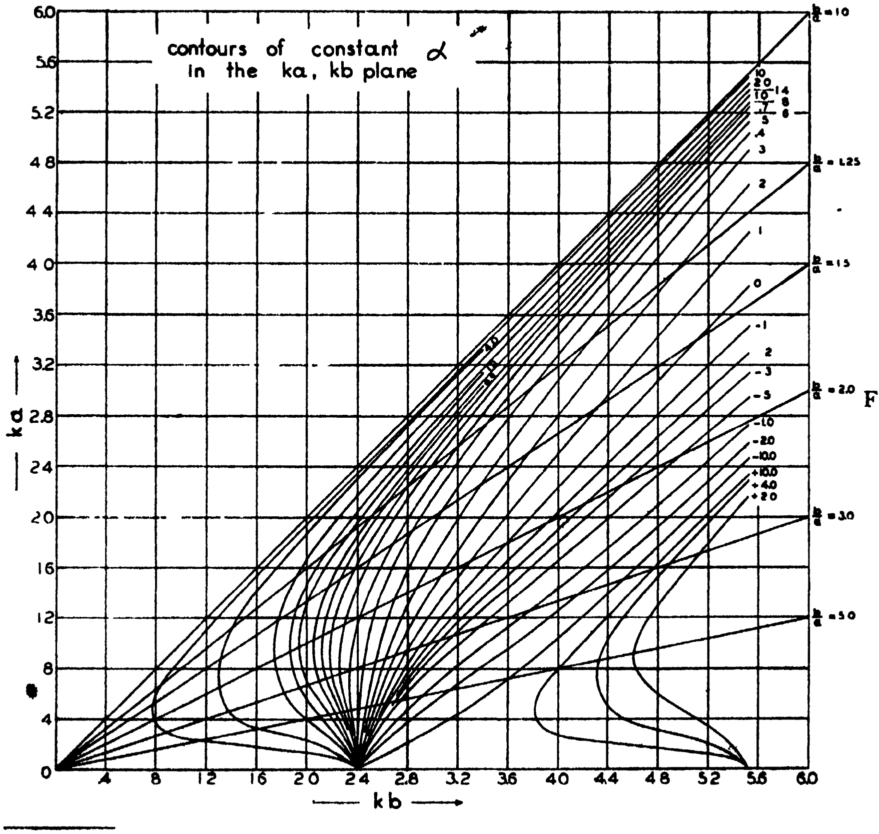


FIG. 5. Shows contours of constant α in the $ka$, $kb$ plane.

---

[1] This notation is similar in spirit, though not in detail, to the standard notation $(x)$ for "the integral part of $x$."

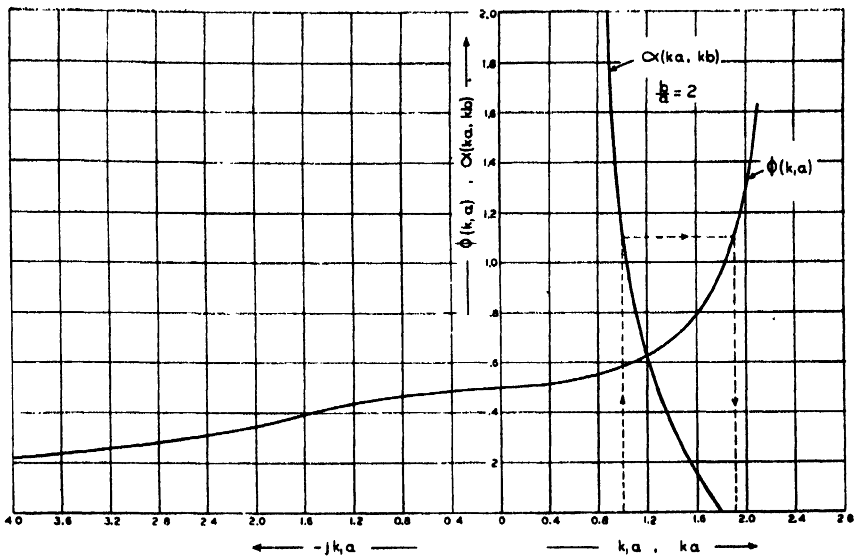


FIG. 6. The functions $\phi$ and $\alpha$ are plotted against $k_1a$ and $ka$ for $b/a = 2$.

center of the cell in which $z$ lies.[2] This function would then be expanded as a Fourier integral for $r<a$ and as a Fourier series for $a<r$, whereupon if the assumed form of $E_z$ is correct, the computed values of $B_\phi$ will match everywhere on $r=a$. To carry such a calculation out in detail would be possible but not simple. In certain cases, however, some progress can be made. In particular, if $d \ll a$ and $d \ll b-a$, a good guess for $f$ will be $[1-(4z^2/d^2)]^{-\frac{1}{2}}$. If we take this form for $f$ without verification we need match only at one point, say at the center of a cell, and if we furthermore carry only one Fourier term in the magnetic-field series for both $r<a$ and $a<r$, we get a crude but tractable result, namely, that one merely multiplies the left-hand side of (5) by $J_0(k_3d/2)$ to get

$$J_0(k_3d/2)\cdot\phi(k, a) = \alpha(ka, kb)/1-\eta. \tag{6}$$

This correction, while reaching a maximum value of $J_0(\pi/2)=0.571$ at $k_3d=\pi$, rises slowly and for small values of $k_3d$ is often negligible. In much of what follows, it will be omitted in the interests of simplicity.

To solve this equation, we need values of $\phi$ and $\alpha$. In Fig. 4 we give a plot of $\phi$ for both real and imagihary values of $k_1a$, together with various useful and easily derived approximate formulae. And in Fig. 5 we give a contour map of $\alpha$ in the $ka$, $kb$ plane.

In obtaining this map, and for various future purposes, the following easily derived approximate expressions are useful

$$\begin{aligned}
&\alpha \cong (1/k^2a^2)(1/\ln\xi), \quad \xi=b/a. \quad ka \ll 1, \quad kb \ll 1,\\
&\alpha \cong (1/k^2a^2)[1/\ln(2/\gamma ka)], \quad \gamma=1.781,\\
&\qquad\qquad ka \ll 1, \quad J_0(kb)\neq 0,\\
&kb \cong 2.405+1.545(1/2-\alpha)k^2a^2,\\
&\qquad\qquad ka \ll 1, \quad kb \cong 2.405,\\
&\alpha=(1/ka)\cot(kb-ka), \quad 1 \ll ka, \quad 1 \ll kb.
\end{aligned} \tag{7}$$

For given $b$ and $a$, and therefore fixed $b/a=\xi$, $\alpha$ is now a function of $k$ or of $ka$, which can be determined from Fig. 5. A typical curve of $\alpha$ *vs.* $ka$, for $kb/ka=2$ is plotted in Fig. 6. On this same graph, we have plotted $\phi$ as a function of $k_1a$.

To solve (5) for given $k$ and $\eta=0$, we enter the chart at $ka$ and follow the vertical dotted line to an intersection with the $\alpha$-curve, proceed horizontally to find an equal value of $\phi$, and then drop down to the axis to find the value of $k_1a$. Finally, we compute $(k_3/k)=(1-(k_1a/ka)^2)^{\frac{1}{2}}$ and so are able to construct a graph of $k_3/k$ *vs.* $ka$ or $kb$.

It is instructive to follow the results qualitatively as $ka$ increases from zero. For $ka$ small, $\alpha$ and so $\phi$ are large and we easily find $(k_3/k) \cong j(2.405/ka)(1-(ka/2.40)^2)^{\frac{1}{2}}$, i.e., the guide acts

---

[2] A justification of this statement would be simply a paraphrase of Floquet's theorem on the solutions of Hill's equation. See, for example, Whittaker and Watson, *Modern Analysis* (Cambridge University Press, New York, 1945), p. 412.

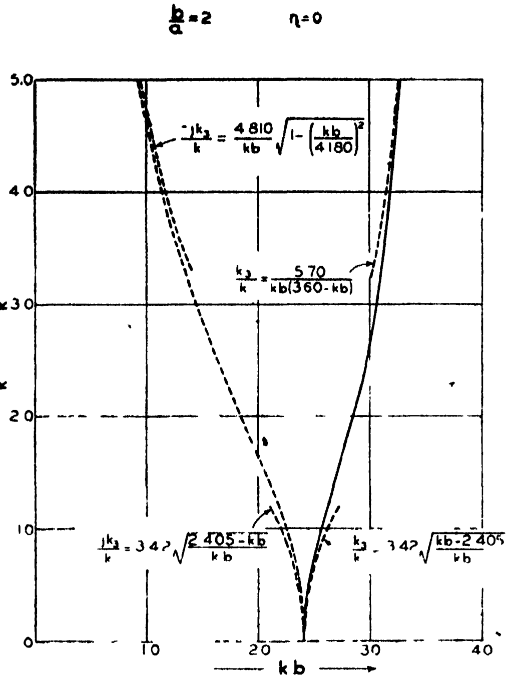


FIG. 7. Shows $k_3/k$ as a function of $kb$ for $b/a=2$, $\eta=0$, $kd\cong0$. Also shown are a number of approximate formulae.

as an attenuator, and with the attenuation characteristic of a tube of radius $a$. The wave number $k_3$ first becomes real when $k_1a=ka$, i.e., at the intersection of the two curves, where $\alpha$ and $\phi$ are usually about 0.5. Moreover, at this low frequency cut-off, which occurs at $kb=2.405$, $k_1$ is a linear function of $k$ and so $k_3/k$ has the approximate form $(k_3/k)\cong(2(kb-2.405)/2.405)^{\frac{1}{2}}\cdot(-dk_1/dk)^{\frac{1}{2}}$, and so the curve has a vertical tangent at cut-off. As we further increase $ka$, $k_1a$ decreases and the next interesting point is at $k_1a=0$, where $k_1a$ changes from real to imaginary, and $k_3/k$ passes through unity. Nothing striking happens to $k_3/k$, it being easily verified that its derivative is given by

$$\frac{dk_3}{dk}=\frac{k}{k_3}\,\frac{k_1}{k_3}\,\frac{d\alpha/dka}{d\phi/dk_1a}\,\frac{1}{1-\eta},\qquad k_1a\neq0, \tag{8}$$

and

$$\frac{dk_3}{dk}=1-\frac{1}{1-\eta}\,\frac{8}{ka}\,\frac{d\alpha}{dka},\qquad k_1a\cong0. \tag{9}$$

These formulae are of some importance, for they give the reciprocal of the group velocity, and this quantity is often of direct interest, especially in the special case $k_1=0$, $k_3=k$ where the wave velocity is $c$.

As $ka$ increases further, $\alpha$ and so $\phi$ approach zero, $k_1$ becomes large and imaginary, and $k_3$ large and real. An approximate formula, valid in

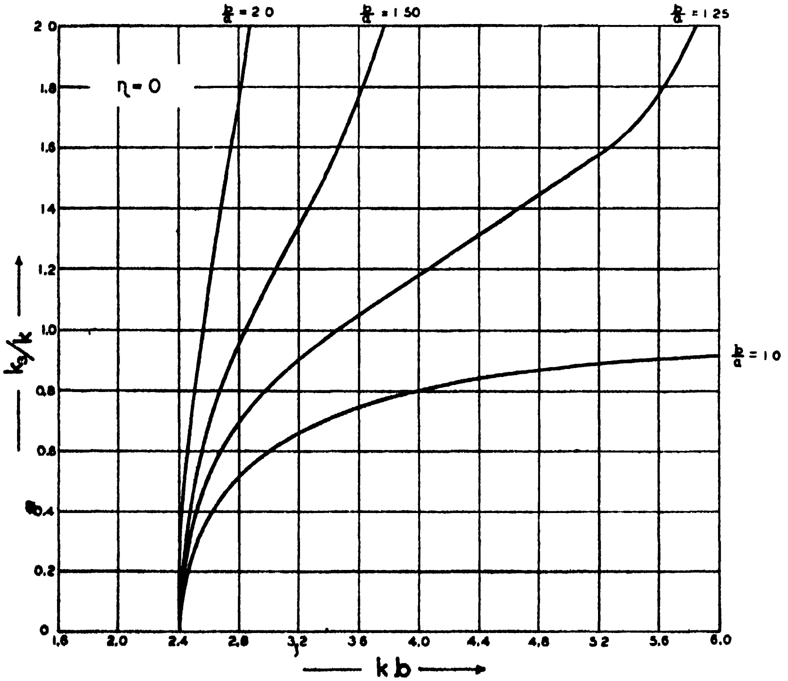


FIG. 8. Shows $k_3/k$ as a function of $kb$ for a variety of ratios $b/a$.

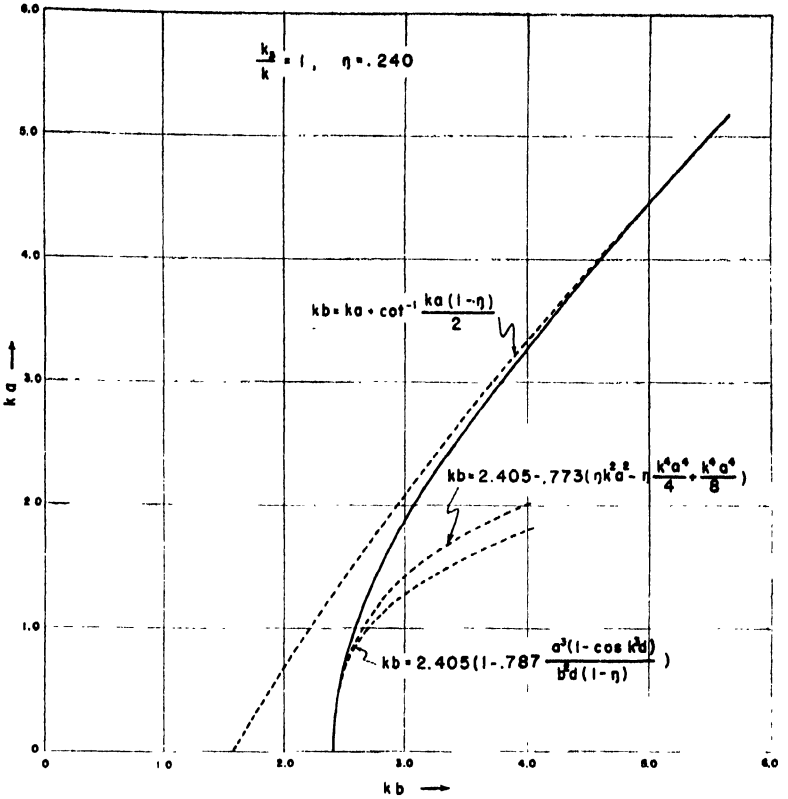


FIG. 9. The value of $ka$ that makes $k_3=k$ is plotted against $kb$. The full curve is based on the assumption that $d \ll a$, $d \ll b-a$, as are two of the dotted curves which give approximations to the solid curve. Also shown dotted is a result valid in the limit $a \ll d$, $a \ll b$.

this region, is

$$(k_3/k)=0.571((1-\eta)/ka(ka-k^+a))dka/d\alpha,$$

with $k^+a$ the value of $ka$ which makes $\alpha=0$.

As remarked before, the present theory is not good in this region, for when $k_3d=\pi$, the wave functions are not well approximated by single Fourier terms. Nevertheless, a good guess at the high frequency cut-off may be had by assuming it to occur at $k_3d=\pi$.

To illustrate the above quantitatively we have prepared Fig. 7 which shows $k_3/k$ as a function of $kb$ for $kb/ka=2$, $\eta=0$, together with a number of approximate formulae. In Fig. 8, $k_3/k$ is plotted as a function of $kb$ for several values of $kb/ka$. Both these figures assume that $k_3d$ is small.

Guides operated at $k$ and loading values such as to give $k_3/k=1$, i.e., phase velocity equal to that of light, have been of special interest to us, and in Fig. 9 we plot for various values of $kb$ the value of $ka$ that gives $k_3/k=1$. Two approximations are also shown, namely,

$$kb=2.405+0.772[\eta k^2a^2-\eta(k^4a^4/4)+(k^4a^4/8)], \quad ka \ll 1, \quad kd \ll 1, \quad (10)$$

$$kb=ka+\cot^{-1}(ka/2)(1-\eta), \quad 1 \ll ka, \quad kd \ll 1/ka. \quad (11)$$

We may note that as $ka$ approaches zero it must inevitably become smaller than $kd$, in which case our analysis breaks down. Nevertheless, if $d \ll b$ there is a region in which (10) is useful.

We may now develop the theory for the case $a \ll d$. Here we proceed by considering the loaded tube as derived from a sequence of uncoupled cavities perturbed by the introduction of coupling holes. The problem is solved by variational methods, the behavior of the trial function near the coupling hole being approximated by finding a solution of the appropriate static problem. The result is

$$kb \cong 2.405\left[1+0.787\frac{a^3}{b^2d(1-\eta)}(1-\cos k_3d)\right], \quad (12)$$
$$\eta d \ll a \ll d, \quad a \ll \lambda,$$

and this result is also plotted in Fig. 9.

On the other end of the possible range of values of $a$ we may have $b-a \ll d$, and in this

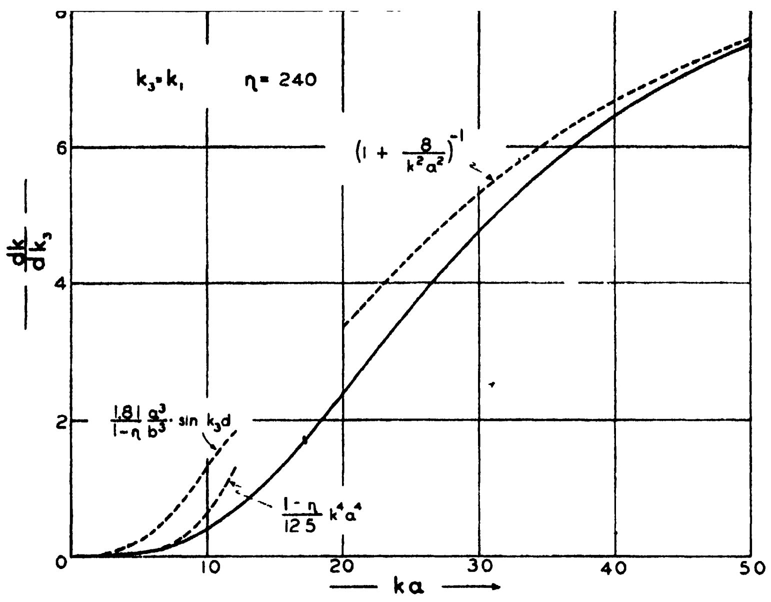


FIG. 10. The quantity $dk/dk_3$, which is the group velocity divided by $c$, is shown as a function of $ka$. The solid line is computed from a theory valid when $d \ll a$, $d \ll b-a$, as are two approximate results shown dotted. Also shown is the result of a theory valid when $a \ll d$, $a \ll b$.

case it is obviously best to take as the unperturbed problem an unloaded guide and introduce the loading disks as perturbations. There are then two cases, $(b-a) \ll \eta d$ and $\eta d \ll b-a \ll d$.

In the first case we may take as a perturbed field simply the unperturbed fields with the volume occupied by the loading rings removed. The results are most easily expressed by giving $k$ in terms of the various dimensions and $k_3$, thus

$$k = k_3[1+(2.405/k_3b)^2]^{\frac{1}{2}} \times[1+5.77(\eta k[b-a]/[kb]^2)], \quad (b-a) \ll \eta d. \quad (13)$$

In the other case, when $\eta d < b-a < d$, the loading disk may be considered as thin, and we use as a perturbed field the unperturbed one minus a potential function such that the boundary conditions on $E$ are satisfied over the ring while $E$ is undisturbed at large distances from the ring. This function is found by a solution of Laplace's equation, under the simplifying assumption that the cylindrical problem may be unrolled into a plane one, i.e., $b-a \ll b$. From this electric field a corresponding magnetic field is derived and the change in frequency computed in terms of the surface integral of $E_1 \times B_2$, with $E_1$ the original field and $B_2$ an addition due to the perturbation. The result is, then

$$k = k_3[1+(2.405/k_3b)^2]^{\frac{1}{2}} \times[1-\pi/2(k_3/k)^2([b-a]^2/bd)]. \quad (14)$$

The group velocity for guides loaded to give a phase velocity of $c$ has also been of interest, and in Fig. 10 we plot $dk/dk_3$ against $ka$ for $kb = 2.66$ together with various approximate formulae. One of these approximate formulae, good for $a \ll d$, is obtained simply by differentiating (12), the other two are derived on a basis best explained in a following section.

### 2. Impedance, Power Flow, and Energy Storage

As is usual in distributed constant problems, any number of definitions of characteristic impedance are possible: the one that is useful for connecting guides of different impedances is defined as equal to the ratio of the transverse voltage to the wall current. If we evaluate this in the midplane of a loading disk using the fields of (1) and (2) we find the impedance in e.m.u. to be

$$Z_t = c\frac{k_3}{k}\frac{2(1-J_0(k_1a))}{k_1aJ_1(k_1a)}. \quad (15)$$

For many purposes it is more useful to know the power flow, and this can be expressed either in terms of $E_0$, the peak axial field, or in terms of the integral of $E_r$, designated above as the transverse voltage. The power flow is easily evaluated from the Poynting vector, the result being

$$P=(1/256\pi^2)k^3k_3a^4[8/(k_1a)^2] \times [J_1^2(k_1a)-J_0(k_1a)J_2(k_1a)](E_0\lambda)^2/c. \quad (16)$$

A plot of $8/(k_1a)^2$ times the bracketed Bessel functions is given in Fig. 11.

A special case of interest is that for which $k_3=k$, $k_1=0$, in which case we find

$$E_0\lambda=(4/\pi)(Pc)^{\frac{1}{2}}(\lambda/a)^2, \quad k_3=k. \quad (17)$$

If one wishes to use radial voltage, the conversion can be accomplished by using the formula connecting the peak value of $E_r$ and $E_0$,

$$(1/E_0\lambda)\int_0^a E_r dr=(kk_3/2\pi k_1^2)(1-J_0(k_1a)). \quad (18)$$

### 3. Energy Storage

The mean energy stored per unit length is of interest, not only in itself, but also in connection with $Q$ calculations to follow and for use in an alternative method of finding the group velocity. This latter use derives from the relation

$$(\text{Energy stored/cm})(\text{group velocity})=P, \quad (19)$$

and owes its usefulness to the facts that $P$ is easy to compute and that it is often not hard to approximate the energy storage, whereas approximations to $dk/dk_3$ have not been obvious.

We may note that when the group velocity is computed from (19) it comes out in terms of integrals of squares of Bessel functions, whereas when computed from $dk/dk_3$ it depends on derivatives of Bessel functions and it is, at first sight, hard to see in detail why the results are the same. We will not go into this matter, but remark that it is similar to the well-known theorem whereby the derivative of an impedance, with respect to frequency, can be found from the energy stored in the impedance.

We will not write down the exact formulae for energy storage, since the result is easy to obtain analytically and most tedious to use numerically.

Two approximations are useful, however. In

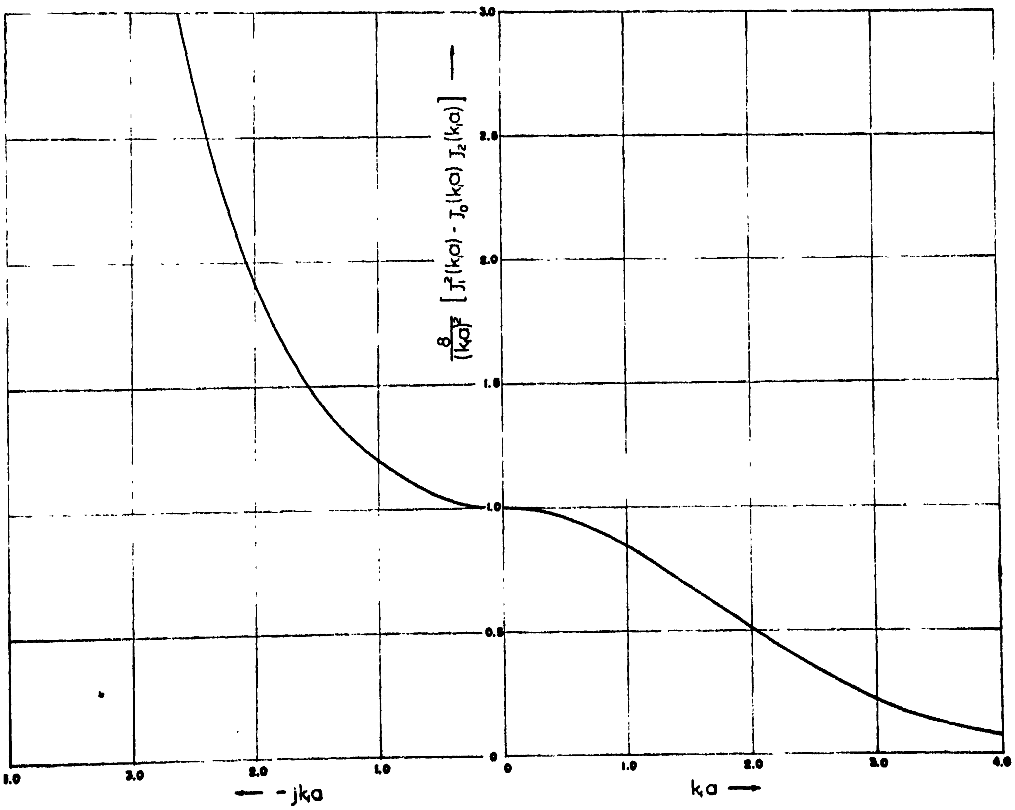


FIG. 11. Plot of a function useful in computing power flow in loaded wave guides.

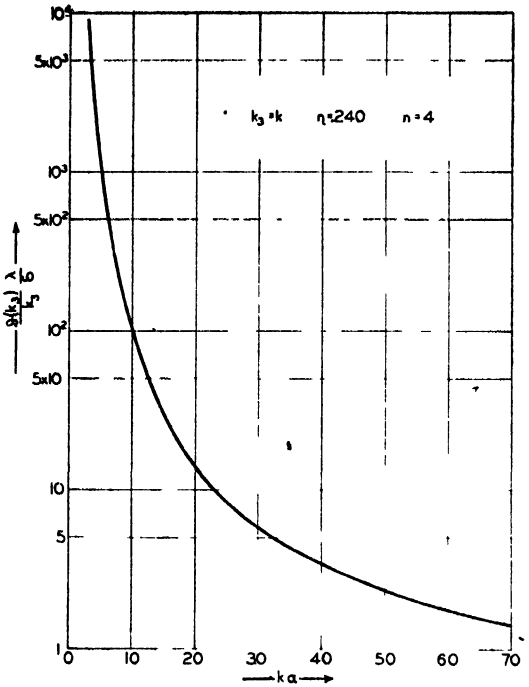


FIG. 12. The imaginary part of the propagation constant, multiplied by $\lambda/k_3\delta$ is plotted against $ka$. The radius $b$ is such as to give $k_3=k$, there are 4 loading disks per wavelength, and the fields of Eqs. (1) and (2) are used

the first place, if $ka\rightarrow 0$ we can approximate $B_\phi$ quite well as $[2.405E_0/(1-\eta)kb]J_1[2.405(r/b)]$ in which case the energy storage, as computed by doubling the magnetic energy, comes out as

$$\text{Mean energy/cm}\cong 0.195E_0^2/(1-\eta)k^2,\qquad ka\ll 1,\tag{20}$$

and combining this with the expression for power flow we get one of the approximations for $dk/dk_3$ used in Fig. 10.

Next, consider the special case $k_3=k$ and let $kb\rightarrow\infty$, so that $kb\cong ka+(2/(1-\eta)ka)$. Then we have $B_\phi=E_0kr/2$ for $r<a$, and we can take $B_\phi\cong(E_0/1-\eta)(1/(kb-ka))=\text{constant}$ for $a<r<b$. This approximation is obviously good for large $ka$ and is surprisingly close for small $ka$, being only 1.55 times too small for $ka=0$. Using these fields, we find

$$\text{Mean energy/cm}\cong\frac{E_0^2}{8k^2}\left(\frac{k^4a^4}{8}+\frac{1}{1-\eta}\,\frac{kb+ka}{kb-ka}\right),\qquad k_3=k,\quad 1\ll ka,\tag{21}$$

and this may be further simplified by using (11) to relate $kb$ and $ka$. Doing so, we get the remaining approximate formula of Fig 10.

## 4. Losses

If a power $P$ crosses some plane transverse to the guide, the power at some distance down the guide will be smaller because of losses, i.e., $dP/dz$ is finite. Before computing the losses, we must first consider how the results are to be expressed. At least three forms are useful. First, we may compute the voltage attenuation $I(k_3)=(1/2P)(dP/dz)$; second, we can find $Q=2\pi\nu$ (energy stored/cm)/$(dP/dz)$; and third, we can compute $(dP/dz)/(E_0\lambda)^2$.

The first two may be related by way of (19), thus

$$Q=(1/2)[k/I(k_3)](dk_3/dk).\tag{22}$$

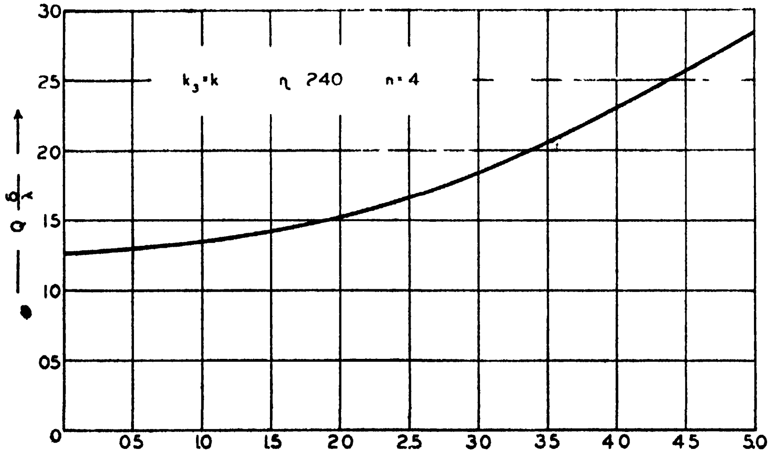


FIG. 13. The quantity $Q\delta/\lambda$ is plotted against $ka$ under the same assumptions as in Fig. 12.

The actual calculation is simple enough in principle, the losses being evaluated as $B^2\delta/8$ per cm² per cycle, with $\delta=(\rho/2\pi\omega)^{\frac{1}{2}}$ the skin depth. The detailed results will not be written down since the formulae, while easy to obtain, are too complex to allow one to determine their behavior by inspection.

The results of a typical calculation are given in Figs. 12, 13, and 14, which give $I(k_3)/k_3(\lambda/\delta)$, $Q(\delta/\lambda)$, and $(\lambda^2/\delta)(dP/dz)/(E_0\lambda)^2$ as functions of $ka$, when $kb$ is such that $k_3=k$ and there are four loading disks per wave-length.

While the general formulae are quite complex, some approximations will now be given which are usefully simple. All of these are for the special case $k_3=k$.

First, let $ka\ll 1$. Then we find, by using the same field as that used for (20),

$$\frac{I(k_3)}{k_3}=\frac{\delta}{\lambda}\frac{6.25}{k^4a^4(1-\eta)^2}[n+2.61(1-\eta)],$$

$$ka\ll 1,\quad k_3=k, \tag{23}$$

$$Q=(\lambda/\delta)[1-\eta/n+2.61(1-\eta)], \tag{24}$$

$$[1/(E_0\lambda)^2](dP/dz)=0.0308(\delta/\lambda^2c)$$
$$\times[1/(1-\eta)^2][n+2.61(1-\eta)], \tag{25}$$

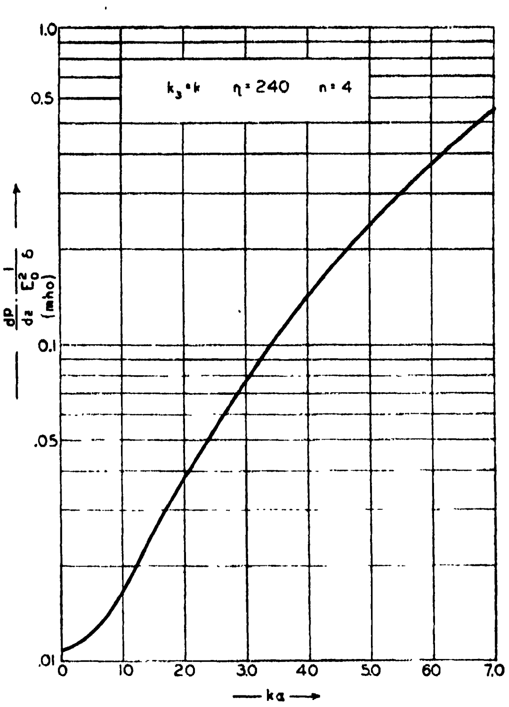


FIG. 14. The power loss per cm per volt/cm squared, divided by $\delta$, as measured in mho, is plotted against $ka$, the conditions being as described for Fig. 12.

with $n=2\pi/kd$, the number of loading disks per wave-length.

Next, let $1\ll ka$ and use the approximation of (21), when we find

$$\frac{I(k_3)}{k_3}\cong\frac{\delta}{\lambda}\frac{8\pi}{k^4a^4}\left[\frac{kb}{(1-\eta)(kb-ka)^2}+\frac{n}{2\pi(1-\eta)^2}\frac{kb+ka}{kb-ka}+\frac{n}{2\pi}\frac{\eta k^2ad}{(1-\eta)^2(kb-ka)^2}\right], \tag{26}$$

$$Q\cong\frac{\lambda}{\delta}\left[\frac{k^4a^4}{8}+\frac{1}{1-\eta}\frac{kb+ka}{kb-ka}\right]\Big/\left[\frac{2\pi kb}{(1-\eta)(kb-ka)^2}+\frac{n}{(1-\eta)^2}\frac{kb+ka}{kb-ka}+\frac{n\eta k^2ad}{(1-\eta)^2(kb-ka)^2}\right], \tag{27}$$

$$\frac{1}{(E_0\lambda)^2}\frac{dP}{dz}=\frac{\delta}{\lambda^2c}\frac{1}{8}\left[\frac{kb}{(1-\eta)(kb-ka)^2}+\frac{n}{2\pi(1-\eta)^2}\cdot\frac{kb+ka}{kb-ka}+\frac{n}{2\pi}\frac{\eta k^2ad}{(1-\eta)^2(kb-ka)^2}\right]. \tag{28}$$

When $ka$ is sufficiently large, these may be simplified by using the approximate relation (11), with results

$$I(k_3)/k_3\cong(\lambda/b)(\delta/\lambda), \tag{29}$$

$$Q\cong(b/2\lambda)(\lambda/\delta), \tag{30}$$

$$\frac{1}{(E_0\lambda)^2}\frac{dP}{dz}\cong\frac{1}{32}\frac{k^3b^3\delta}{\lambda^2c}. \tag{31}$$

A qualitative understanding of the above formulae is useful and simple. For $k_3=k$ we have $E_r=B_\phi=E_0kr/2$ for $r<a$, and we will not do too badly by continuing $B_\phi$ in a linear manner to $r=b$. Thus the power flow will be proportional to $k^4a^4$, the energy stored to $k^4b^4$ and the energy loss on the walls will vary like $k^3b^3$, all for constant axial field. Thus the group velocity, which depends on the ratio of power transmitted to energy stored, at first rises rapidly as $ka$ increases and

then levels off as $ka$ approaches $kb$. Likewise, the linear attenuation, which depends on the ratio of power loss to power transmission, starts by decreasing rapidly as $ka$ increases. When $ka$ is large, the decrease is slower, being simply due to greater cross-sectional area for the wave to carry power as compared to the perimeter bounding the wave and introducing losses. On the other hand, $Q$ which depends on the ratio of energy stored to energy loss, depends mainly on $kb$ which at first increases only slowly with rising $ka$. Thus $Q$ at first drops slowly, though finally varying like $1/ka$, for essentially the same reason as the attenuation.

## CONCLUSION

Many of the important properties of disk-loaded wave guides have been developed, both qualitatively and quantitatively, and have been described by formulae, both approximate and exact, and a number of curves, which will, it is hoped, illustrate the more important points. Special attention has been paid to the wave number $k_3$ which has been computed by three methods. On the other hand, the formulae for loss and power flow are, for the most part, based on an approximation valid when the disk spacing is small, i.e., the loading is nearly continuous.

In a problem with so many variables, it is impossible, in a finite article, to cover all the proportions and conditions that might be of interest and we therefore do not feel apologetic when we point out some of the gaps in the present treatment. First, more accurate methods of calculating $k_3$ are certainly possible; the virtue of our methods is that they give closed or nearly closed formulae and so can be used to explore a wide range of variables without undue labor. Second, we have hardly touched on the behavior in the vicinity of the high frequency cut-off. Third, the loss and power-transmission calculations have been restricted to one method. And fourth, many of the illustrative graphs have been for a single case. Despite these lacunae, we hope that our results will be of use to others.

Finally, it is a pleasure to record the assistance of Mr. E. T. Jaynes with many of the early calculations and to acknowledge the liberal help of the Navy under Contract N6-ori-106.

# Equations for the Inductances of Three-Phase Coaxial Busses Comprised of Square Tubular Conductors

THOMAS JAMES HIGGINS
*Illinois Institute of Technology, Chicago, Illinois*

AND

HENRY PETER MESSINGER
*University of Illinois, Urbana, Illinois*

(Received July 15, 1947)

Equations are derived for calculating the associated inductances of the conductors of three-phase coaxial busses comprised of square tubular conductors. It is assumed that the conductors are nonmagnetic, are of such lengths that end effects are negligible, are right-cornered, and carry currents uniformly distributed over their cross sections. The general equations are obtained through use of geometric mean distance theory. Approximate equations, the conductors being considered as indefinitely thin, are epitomized in a table which yields values sufficiently accurate for most design work. In turn, these approximate equations are reduced to simple equations which also yield values sufficiently accurate for much design work. These equations for inductance lead to corresponding equations for the reactive voltage drops. Use of these various equations and the relative accuracy to be expected of them is illustrated by calculating the inductances and reactive voltage drops of typical busses.

The changes in the inductances due to rounding the edges of the conductors and the changes produced in both the inductances and the resistances by skin and proximity effects are investigated. The changes in the inductances due to rounding the edges can be approximated by use of a known equation. The changes in the inductances due to power frequency skin and proximity effects prove to be negligible; the changes in the resistances can be calculated by use of known skin-effect factors and newly derived proximity-effect factors.

Calculation of the inductances of three-phase coaxial busses constructed of tubular conductors made up of two channels or angles placed flange to flange is discussed. Finally, attention is called to a relatively little known means of obtaining balanced three-phase operation of geometrically unbalanced three-phase busses.

BY virtue of its axial-symmetric geometry, the coaxial-tubular bus possesses certain very desirable electrical and mechanical characteristics. Low inductance contributes to good voltage regulation. Reduced skin and proximity effects minimize line losses and average temperature rise under load. Axial-symmetric distribution of skin and proximity effects reduces current density distortion and attendant hot-spot concentrations, hence enables a greater current rating for a specified maximum copper temperature. As electromagnetic forces between conductors are largely eliminated, mechanical bracing against short-circuit forces can be correspondingly reduced. The nested conductors require less space for installation and a smaller supporting structure than they would if arranged side by side.

Yet, despite these desirable characteristics, the paucity of published descriptions of actual installations[1–4] indicates that few coaxial busses have been constructed to date. However, until the last few years factory distribution loads at frequencies greater than 60 cycles were not common. Again, most 60-cycle loads were supplied over relatively short distances. As these conditions are such that satisfactory operation can be obtained with the more conventional, though less electrically effective, construction of conductors placed side by side, this limited use of coaxial busses is not surprising. Now, however, by virtue of manufacturing procedures and techniques developed during the war, the bus designer is faced with the necessity of designing heavy-distribution busses which are not only long but also are to operate at frequencies of several hundred cycles and higher. But, the electrical characteristics of the coaxial bus especially suit it to satisfactory operation under either or

---

[1] L. R. Bogardus, "Resistance welder feed has low resistance drop," Elec. World 110, 702–704 (1938).

[2] W. Specht, "Concentric bus for resistance welders," Elec. World 114, 558 (1940).

[3] W. Specht, "A concentric channel bus for resistance welders," Weld. Eng. 26, June, 31–32 (1941).

[4] "Power supply for resistance welding machines," Trans. A.I.E.E. 59, 306–320 (1940); 60, 185–192 (1941).

both of these conditions. Hence, it is to be expected that in the near future coaxial busses will find greater use than they have hitherto.

At present coaxial busses are constructed of circular tubular conductor, of square tubular conductor, and of structural tubular conductor comprised of two slightly separated channels or angles placed flange to flange. To predetermine the operating characteristics of these busses requires knowledge of the a.c. resistance, the a.c. inductance, the copper temperature under load, and the mechanical forces exerted during short-circuit. Equations[5,6] for calculating all four of these quantities are available for single or three-phase busses constructed of circular-tubular conductor and for single-phase busses constructed of square or structural tubular conductor. But for three-phase busses constructed of square or structural tubular conductor, though equations are available for calculating the a.c. resistance,[7,8] the copper temperature,[8] and the short-circuit force,[9] equations for calculating the a.c. inductance, *usually the determining factor in the design of a bus*, are not available.

Such equations are derived in this paper, it being assumed that the conductors are non-magnetic, of such lengths that end effects are negligible, right-cornered, and carry currents distributed uniformly over their cross sections. Of these four conditions, the first two are satisfied in practice. The third, however, though essential to tractable analysis, is usually not true in practice. Square tubes, channels and angles, drawn, rolled, bent, or otherwise formed in one piece, have rounded edges. But the error introduced by neglecting the rounded edges is small. Commonly, the result obtained, though a bit low, is yet sufficiently accurate for the work in hand. In any case, the incremental inductance associated with the rounded edges can be approximated closely by a simple semi-empirical equation derived by Dwight and Wang.[10] Finally, though skin and proximity effects render the current density over the cross sections of the conductors slightly non-uniform, detailed analysis[6] indicates and experiment confirms[11] that at power frequencies the a.c. inductance is practically identical with the d.c. inductance.

These four postulates granted as consistent with practice, derivation of the desired equations, though lengthy, is not difficult. The conductor cross sections are considered as combinations of parallel-sided square areas. By virtue of geometric-mean-distance theory the inductance associated with a particular conductor is expressed in terms of the self and mutual geometric mean distances of these areas. These geometric mean distances can be calculated from known equations. Accordingly, determination of the inductance as a function of the parameters defining the geometry of the bus reduces, essentially, to specific determination of the geometric mean distances involved and substitution of them in the general equation for the inductance.

The general equation comprises 48 terms. A simpler approximate equation similarly derived, but on the assumption that the conductors can be considered as being indefinitely thin (hence have cross sections comprised of line segments), affords values sufficiently accurate for many design purposes. As a further convenience to numerical computation, this equation is expressed in terms of a parameter whose values are to be ascertained from a given table of values. Finally, over the range commonly encountered in practice this parameter proves to be substantially constant. For this case the approximate equation reduces to absurdly simple equations that yield the inductance with negligible computation. These various equations lead to corresponding equations for the reactive voltage drops.

The desired equations, having been obtained

---

[5] T. J. Higgins, "The design of bus-bar industrial distribution systems: an epitomization of available data," Trans. A.I.E.E. 64, 385–400, 486–488 (1945).

[6] T. J. Higgins, "Formulas for the inductance of coaxial busses comprised of square tubular conductors," Trans. A.I.E.E. 65, 328–336 (1946).

[7] A. H. M. Arnold, "The alternating-current resistance of hollow square conductors," J. Inst. Elec. Eng. 82, 537–545 (1938).

[8] H. B. Dwight, G. W. Andrew, and H. W. Tileston, "Temperature rise of bus bars," Gen. Elec. Rev. 43, 213–216 (1940).

[9] T. J. Higgins, "Formulas for calculating short-circuit forces between conductors of structural shape," Trans. A.I.E.E. 62, 659–663 (1943).

[10] H. B. Dwight and T. K. Wang, "Reactance of square tubular bus bars," Trans. A.I.E.E. 57, 762–765 (1938).

[11] H. P. Messinger, *Equations for the Inductance of Coaxial Busses Constructed of Square Tubular Conductor* (Thesis, Library, Illinois Institute of Technology, Chicago, Illinois).

and checked, are used as illustrated by calculation of the inductances of typical busses constructed of square tubular conductor. Changes in inductance and resistance resulting from rounded corners on the conductors and from skin and proximity effects are discussed in detail. Procedures for calculating the inductance of coaxial busses constructed of channels or angles are advanced. Finally, attention is directed to a relatively little known means of obtaining balanced three-phase operation of geometrically unbalanced three-phase busses.

## I. DERIVATION OF THE GENERAL EQUATION

Consider a three-phase bus comprised of three coaxial square tubular conductors (Fig. 1). From well-known theory[12, 13] the inductances associated with the conductors are

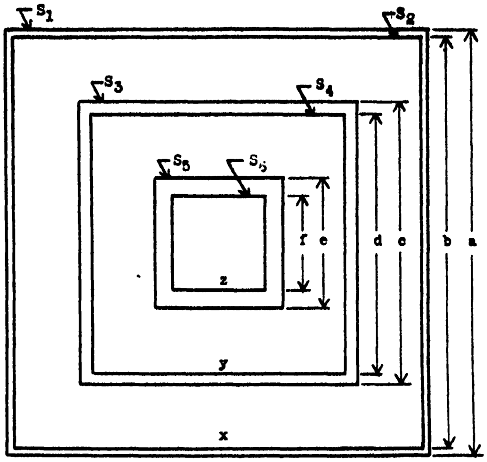


FIG. 1. Cross section of coaxial bus comprised of square tubular conductors.

$$L_x=(-2/I_x)\sum_{i=x}^{z} I_i \log D_{xi}, \tag{1}$$

$$L_y=(-2/I_y)\sum_{i=x}^{z} I_i \log D_{yi}, \tag{2}$$

$$L_z=(-2/I_z)\sum_{i=x}^{z} I_i \log D_{zi}, \tag{3}$$

where, typically, $L_x$ is the inductance associated with conductor $x$; $I_x$ is the current (expressed as a complex number) associated with conductor $x$; and $D_{xi}\equiv D_{ix}$ is the geometric-mean distance of the cross sections of conductors $x$ and $i$. The units of the various physical quantities are as discussed in Section V.

The geometric mean distances $D_{xx}$, $D_{xy}$ and $D_{xz}$ are (as derived in detail in Appendix I)

$$\begin{aligned}\log D_{xx}=&[1/12(a^2-b^2)^2][12a^4\log a+12b^4\log b+(a^4-6a^2b^2+b^4)\log(a^2+b^2)-(a+b)^4\log(a+b)\\&-(a-b)^4\log(a-b)+8ab(a^2-b^2)\tan^{-1}(a-b)/(a+b)+(a+b)(a-b)^3\pi\\&+10(a^4+b^4)\log 2]-0.5\log 2+\pi/4-25/12,\end{aligned}\tag{4}$$

$$\begin{aligned}\log D_{xy}=&[-1/24(a^2-b^2)(c^2-d^2)][(a+d)^4\log(a+d)+(a-d)^4\log(a-d)-(a^4-6a^2d^2+d^4)\\&\times\log(a^2+d^2)-(b+d)^4\log(b+d)-(b-d)^4\log(b-d)+(b^4-6b^2d^2+d^4)\log(b^2+d^2)\\&-(a+c)^4\log(a+c)-(a-c)^4\log(a-c)+(a^4-6a^2c^2+c^4)\log(a^2+c^2)+(b+c)^4\log(b+c)\\&+(b-c)^4\log(b-c)-(b^4-6b^2c^2+c^4)\log(b^2+c^2)-8ad(a^2-d^2)\tan^{-1}(a-d)/(a+d)\\&+8bd(b^2-d^2)\tan^{-1}(b-d)/(b+d)+8ac(a^2-c^2)\tan^{-1}(a-c)/(a+c)-8bc(b^2-c^2)\\&\times\tan^{-1}(b-c)/(b+c)-\{(a+d)(a-d)^3-(b+d)(b-d)^3-(a+c)(a-c)^3\\&+(b+c)(b-c)^3\}\pi]-0.5\log 2+\pi/4-25/12,\end{aligned}\tag{5}$$

$$\begin{aligned}\log D_{xz}=&[-1/24(a^2-b^2)(e^2-f^2)][(a+f)^4\log(a+f)+(a-f)^4\log(a-f)^4-(a^4-6a^2f^2+f^4)\\&\times\log(a^2+f^2)-(b+f)^4\log(b+f)-(b-f)^4\log(b-f)+(b^4-6b^2f^2+f^4)\log(b^2+f^2)\\&-(a+e)^4\log(a+e)-(a-e)^4\log(a-e)+(a^4-6a^2e^2+e^4)\log(a^2+e^2)\\&+(b+e)^4\log(b+e)+(b-e)^4\log(b-e)-(b^4-6b^2e^2+e^4)\log(b^2+e^2)-8af(a^2-f^2)\\&\times\tan^{-1}(a-f)/(a+f)+8bf(b^2-f^2)\tan^{-1}(b-f)/(b+f)+8ae(a^2-e^2)\tan^{-1}(a-e)/(a+e)\\&-8be(b^2-e^2)\tan^{-1}(b-e)/(b+e)-\{(a+f)(a-f)^3-(b+f)(b-f)^3-(a+e)(a-e)^3\\&+(b+e)(b-e)^3\}\pi]-0.5\log 2+\pi/4-25/12.\end{aligned}\tag{6}$$

---

[12] T. J. Higgins, "Formulas for the calculation of the inductance of linear conductors of structural shape," Trans. A.I.E.E. 62, 53–58 (1943).

[13] L. F. Woodruff, *Principles of Electric Power Transmission* (John Wiley and Sons, New York, 1939), edition 2, Chap. 2.

Substituting from Eqs. (4)–(6) in Eq. (1), collecting terms, and utilizing the relationship $I_x+I_y+I_z=0$ yields

$$\begin{aligned}
L_x=&[-1/6(a^2-b^2)^2][12a^4\log a+12b^4\log b+(a^4-6a^2b^2+b^4)\log(a^2+b^2)-(a+b)^4\log(a+b)\\
&-(a-b)^4\log(a-b)+8ab(a^2-b^2)\tan^{-1}(a-b)/(a+b)+(a+b)(a-b)^3\pi+10(a^4+b^4)\log 2]\\
&+(I_y/I_x)[1/12(a^2-b^2)(c^2-d^2)][(a+d)^4\log(a+d)+(a-d)^4\log(a-d)\\
&-(a^4-6a^2d^2+d^4)\log(a^2+d^2)-(b+d)^4\log(b+d)-(b-d)^4\log(b-d)+(b^4-6b^2d^2+d^4)\\
&\times\log(b^2+d^2)-(a+c)^4\log(a+c)-(a-c)^4\log(a-c)+(a^4-6a^2c^2+c^4)\log(a^2+c^2)\\
&+(b+c)^4\log(b+c)+(b-c)^4\log(b-c)-(b^4-6b^2c^2+c^4)\log(b^2+c^2)-8ad(a^2-d^2)\\
&\times\tan^{-1}(a-d)/(a+d)+8bd(b^2-d^2)\tan^{-1}(b-d)/(b+d)+8ac(a^2-c^2)\tan^{-1}(a-c)/(a+c)\\
&-8bc(b^2-c^2)\tan^{-1}(b-c)/(b+c)-\{(a+d)(a-d)^3-(b+d)(b-d)^3-(a+c)(a-c)^3\\
&+(b+c)(b-c)^3\}\pi]+(I_z/I_x)[1/12(a^2-b^2)(e^2-f^2)][(a+f)^4\log(a+f)+(a-f)^4\\
&\times\log(a-f)^4-(a^4-6a^2f^2+f^4)\log(a^2+f^2)-(b+f)^4\log(b+f)-(b-f)^4\log(b-f)\\
&+(b^4-6b^2f^2+f^4)\log(b^2+f^2)-(a+e)^4\log(a+e)-(a-e)^4\log(a-e)+(a^4-6a^2e^2+e^4)\\
&\times\log(a^2+e^2)+(b+e)^4\log(b+e)+(b-e)^4\log(b-e)-(b^4-6b^2e^2+e^4)\log(b^2+e^2)\\
&-8af(a^2-f^2)\tan^{-1}(a-f)/(a+f)+8bf(b^2-f^2)\tan^{-1}(b-f)/(b+f)+8ae(a^2-e^2)\\
&\times\tan^{-1}(a-e)/(a+e)-8be(b^2-e^2)\tan^{-1}(b-e)/(b+e)-\{(a+f)(a-f)^3-(b+f)(b-f)^3\\
&-(a+e)(a-e)^3+(b+e)(b-e)^3\}\pi]. \qquad (7)
\end{aligned}$$

This is the desired equation for the inductance associated with conductor $x$. Comparison of Eqs. (1) and (2) indicates that the corresponding equation for $L_y$ follows from Eq. 7 by interchanging $x$ with $y$, $a$ with $c$, and $b$ with $d$. Similarly, comparison of equations 1 and 3 indicates that the corresponding equation for $L_z$ follows from Eq. 7 by interchanging $x$ with $z$, $a$ with $e$, and $b$ with $f$.

## II. CORROBORATION OF THE GENERAL EQUATION

Consider a three-phase bus comprised of conductors $x$, $y$, $z$ carrying currents $I_x$, $I_y$ and $I_z$ which satisfy the relationship $I_x+I_y+I_z=0$. One conductor, say $x$, can be considered as carrying current $I_x$ and the two remaining conductors can be considered as connected in parallel and carrying the return current $(I_y+I_z)$. In effect, therefore, the bus is a single-phase bus with going conductor $x$ carrying current $I_x$ and subdivided return conductor comprised of parallel components $y$ and $z$ carrying current $(I_y+I_z)=-I_x$. The inductance of such a bus is

$$L=L_x+L_yL_z/(L_y+L_z). \qquad (8)$$

If now conductors $y$ and $z$ are considered as identical and superimposed (whence $y\equiv z$), the resulting single-phase bus is, effectively, comprised of two coaxial square tubular conductors. Accordingly, if Eq. 7 is correct, Eq. (8) should yield the known equation of such a bus. By virtue of $y\equiv z$, we have $e\equiv c$ and $f\equiv d$. Utilizing these identities in Eq. (7) for $L_x$ and in an analogous equation for $L_y\equiv L_z$, substituting the resulting expressions in Eq. (8), and collecting terms yields, corroboratively, the known solution (Eq. 9 of reference 6) for a single-phase bus comprised of two coaxial square tubular conductors.

## III. APPROXIMATIONS TO THE GENERAL EQUATION

The general equation for the inductance associated with a conductor comprises 48 terms: 29 logarithmic, 9 arctangentic, and 10 algebraic. But by use of the proper tools (an electric calculating machine and the 15-place Work Projects Administration[14, 15] tables of natural logarithms and arctangents), the computation incident to the calculation of the associated inductances of a given bus can be effected without difficulty. The computation, however, will be lengthy (the computation incident to example 1 below required approximately 15 hours to perform on a 10-row electric calculating machine by one of the authors, experienced in such computation). For this reason it is most desirable to have approximate equations which enable rapid determination of values sufficiently accurate for much (in many

---

[14] *Tables of Natural Logarithms* (Columbia University Press, New York).

[15] *Tables of Arctangents* (Columbia University Press, New York).

instances, all) of the computation incident to the final design of a bus.

A means of deriving such equations stems from the writers' earlier investigation[6] of the inductance of a single-phase coaxial bus comprised of square tubular conductors. Therein, values of inductance calculated from the general equation were found to be in excellent agreement with values of inductance calculated from much simpler approximate equations. These equations are derived on the basis that the cross sections of the conductors are comprised of line segments of dimensions equal to the external dimensions of the actual conductors.

Consider, then, the cross section of each conductor as comprised of four line segments (Fig. 2). As previously, the associated inductances are given by Eqs. (1)–(3). Therein, however, the geometric mean distances are now those pertinent to the cross sections of Fig. 2. The geometric mean distances $D_{xx}$, $D_{xy}$ and $D_{xz}$ are (as derived in detail in Appendix II)

$$\log D_{xx}=\log a+(1/4)\log 2+\pi/4-3/2, \qquad (9)$$

$$\begin{aligned}\log D_{xy}=&(1/8ac)[(a+c)^2\log(a+c)\\&-(a-c)^2\log(a-c)+2ac\log(a^2+c^2)\\&-2(a^2-c^2)\tan^{-1}(a-c)/(a+c)\\&+0.5(a^2-c^2)\pi]-0.5\log 2\\&+\pi/4-3/2, \qquad (10)\end{aligned}$$

$$\begin{aligned}\log D_{xz}=&(1/8ae)[(a+e)^2\log(a+e)\\&-(a-e)^2\log(a-e)+2ae\log(a^2+e^2)\\&-2(a^2-e^2)\tan^{-1}(a-e)/(a+e)\\&+0.5(a^2-e^2)\pi]-0.5\log 2\\&+\pi/4-3/2, \qquad (11)\end{aligned}$$

Substituting from Eqs. (9)–(11) in Eq. 1, collecting terms and utilizing the relationship $I_x+I_y+I_z=0$ yields

$$\begin{aligned}L_x=&-2\log a-(3/2)\log 2\\&-(I_y/I_x)(1/4ac)[(a+c)^2\log(a+c)\\&-(a-c)^2\log(a-c)+2ac\log(a^2+c^2)\\&-2(a^2-c^2)\tan^{-1}(a-c)/(a+c)\\&+0.5(a^2-e^2)\pi]-(I_z/I_x)(1/4ae)\\&\times[(a+e)^2\log(a+e)-(a-e)^2\log(a-e)\\&+2ac\log(a^2+e^2)-2(a^2-e^2)\\&+\tan^{-1}(a-e)/(a+e)+0.5(a^2-e^2)\pi]. \qquad (12)\end{aligned}$$

Correspondingly, equations for $L_y$ can be obtained from Eq. (12) by interchanging $x$ with $y$ and $a$ with $c$. Similarly, corresponding formulas for $L_z$ can be obtained in interchanging $x$ with $z$ and $a$ with $e$.

Simple algebraic manipulation of the expressions for the mutual geometric mean distances yields equations especially convenient to computation. Thus, designate $K(r)$ by

$$\begin{aligned}K(r)=&(1/4)[\log(1+A^2)-2\log(1-A^2)\\&-(1/2r)(1-r)^2\log A\\&+(r^{-1}-r)(\pi/4-\tan^{-1}A)+2\log r\\&+\log 2+\pi-6], \qquad (13)\end{aligned}$$

where $A=(1-r)/(1+r)$ and $r=(1-A)/(1+A)$. A table of values of $A$, $r$ and $K(r)$ is given in Appendix III.

Then Eqs. (9)–(11) can be written in the form

$$\log D_{xx}=K(1)+\log a, \qquad (14)$$

$$\log D_{xy}=K(c/a)+\log a, \qquad (15)$$

and

$$\log D_{xz}\doteq K(e/a)+\log a. \qquad (16)$$

Similarly,

$$\log D_{yy}=K(1)+\log c, \qquad (17)$$

$$\log D_{zz}=K(1)+\log e, \qquad (18)$$

and

$$\log D_{yz}=K(e/c)+\log c. \qquad (19)$$

Substituting as indicated in Eqs. (1)–(3), noting that $K(1)=-0.5413$, and utilizing the identity $I_x+I_y+I_z=0$ in appropriate fashion yields

$$\begin{aligned}L_x=&-2\{-0.5413-K(e/a)\\&+(I_y/I_x)[K(c/a)-K(e/a)]\}, \qquad (20)\end{aligned}$$

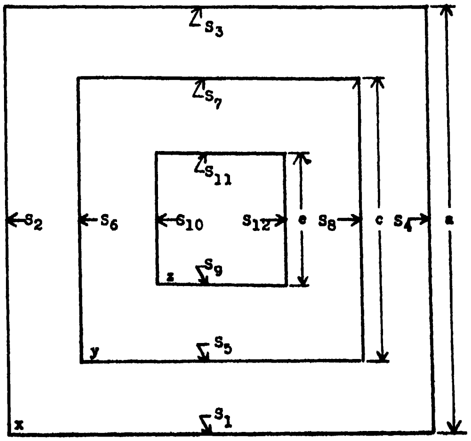


FIG. 2. Cross section of coaxial bus comprised of square tubular line-segment conductors.

$$L_y = -2\{-0.5413 - K(e/c) + (I_x/I_y)\log a/c + (I_x/I_y)[K(c/a) - K(e/c)]\}, \quad (21)$$

$$L_z = -2\{-0.5413 + \log e/c - K(e/c) + (I_x/I_z)\log a/c + (I_x/I_z) \times [K(e/a) - K(e/c)]\}. \quad (22)$$

Now Table I of Appendix III evidences that for $1 \geq A \geq 0.1$, hence $0 \leq r \leq 0.82$, $K(r)$ is substantially constant ranging from $-0.553 \geq K \geq -0.561$. The values of $r$ found in practice approximate the upper limit. Accordingly, selecting $r = 0.75$ as representative, whence

$$K(c/a) \cong K(e/a) \cong K(e/c) \cong -0.556,$$

Eqs. (20)–(22) reduce to the simple approximate equations

$$L_x = -0.030, \quad (23)$$

$$L_y = -0.030 - 2(I_x/I_y)\log a/c, \quad (24)$$

$$L_z = -0.030 - 2\log e/c - 2(I_x/I_z)\log a/c, \quad (25)$$

valid for $c/a$, $e/a$, and $e/c$ not greater than approximately 0.8.

For purposes of computation, the approximate Eqs. (20)–(22) possess considerable advantage over the general Eq. (7). The latter comprises 48 terms, the former only a few. Moreover, these few terms involve the simplest of calculations. The logarithmic term can be read from a slide rule or from standard tables; the values of $K$ from Table I, calculated from Eq. (13). And even this table can be dispensed with in much work. Equations (23)–(25) afford sufficiently accurate values. Thus, in virtue of the fewness and simplicity of the terms comprising the approximate equations, they afford rapid computation.

## IV. EQUATIONS FOR REACTIVE VOLTAGE DROP

In practice, calculation of associated inductance and reactance is incidental to calculation of what is actually desired: namely, the reactive voltage drop on the conductor, equal to $j2\pi fI (= jwI)$. Saving in time of computation results, however, from calculating the reactive voltage drop directly from appropriate equations, to be obtained by multiplying the previously derived equation for indictance by $jwI$. Thus from Eqs. (1)–(3) we have

$$E_x = (-2jw) \sum_{i=x}^{z} I_i \log D_{xi}, \quad (26)$$

$$E_y = (-2jw) \sum_{i=x}^{z} I_i \log D_{yi}, \quad (27)$$

$$E_z = (-2jw) \sum_{i=x}^{z} I_i \log D_{zi}. \quad (28)$$

Particularly useful approximate equations are to be obtained from Eqs. (23)–(25). Thus

$$E_x = -j0.03wI_x, \quad (29)$$

$$E_y = -2jw[0.015I_y + I_x \log a/c], \quad (30)$$

$$E_z = -2jw[I_z(0.015 + \log e/c) + I_x \log a/c]. \quad (31)$$

In similar obvious fashion, corresponding equations can be obtained from Eqs. 7 and analogous equations, from Eq. 12 and analogous equations, and from Eqs. (20)–(22).

## V. UNITS TO BE EMPLOYED

Equations (1)-(3) are valid for the c.g.s. electromagnetic system of units. Hence, the various equations for inductance yield values expressed in abhenrys per centimeter of bus length. Inasmuch, however, as the coefficients and arguments of the various terms comprising these equations can be written in dimensionless form (as, for example, in Eqs. (20)–(25)) we note that in substituting in these equations the currents, and likewise the linear dimensions of the conductors, can be expressed in any like unit.

On the other hand, inasmuch as the voltage drop on a conductor equals $jwLI$, in Eqs. (26)–(31) and in analogous equations the currents must be expressed in abamperes (1 abampere = 10 amperes), though the linear dimensions can yet be expressed in any like unit. The voltage drop, of course, will be in abvolts per centimer of bus length.

For convenience in computation we note the following. Multiplying abhenrys per centimeter by $3.048 \times 10^{-5}$ yields millihenrys per foot of bus length; multiplying by $11.48 \times 10^{-6}$ yields 60-cycle reactance in ohms per foot of bus length. Multiplying abvolts per centimeter by $3.048 \times 10^{-7}$ yields volts per foot of bus length. If in

Eqs. (26)–(31) and in analogous equations the currents are expressed in amperes and the resulting values multiplied by $3.048\times 10^{-8}$ the voltage drops obtained are in volts per foot of bus length. Finally, $\log u(\equiv \log_e u) = 2.3026 \log_{10} u$, where $u$ is an arbitrary number.

## VI. SOME ILLUSTRATIVE EXAMPLES

To illustrate the use of the general and approximate equations and to gain knowledge of the accuracy to be expected of values calculated with the approximate equations, we consider the following examples.

Example 1. The problem is to calculate the six geometric mean distances and the associated inductances of a bus having cross section as in Fig. 1 with dimensions: $a=6$ inches, $b=5.6$ inches, $c=4.5$ inches, $d=4$ inches, $e=3$ inches, $f=1.2$ inches. The bus carries balanced currents $I_x = -0.5-j0.866$, $I_y = -0.5+j0.866$, and $I_z=1$ abampere.

Substituting these values in Eqs. (4)–(6) and analogous equations, rearranging terms, obtaining values of the logarithmic and arctangentic terms from the Work Projects Administration Tables, and effecting all calculations on a 10-row electric calculating machine yields

$$\log D_{xx} = 1.2261, \quad \log D_{xy} = 1.2019,$$
$$\log D_{yy} = 0.9222, \quad \log D_{xz} = 1.2011,$$
$$\log D_{zz} = 0.3576, \quad \log D_{yz} = 0.8733.$$

Substituting in Eqs. (1)–(3) as indicated and collecting real and imaginary terms yields

$L_x = -0.0484$ abhenry per centimeter of bus length,

$L_y = 0.2306-j0.572$ abhenry per centimeter of bus length,

$L_z = 1.359+j0.572$ abhenry per centimeter of bus length.

In calculating these values all logarithmic and arctangentic terms were taken to nine decimal places; values stemming from consequent arithmetic manipulations were held to the same number of decimals. Inasmuch as each of these final values of inductance stems from algebraic addition of some 50 terms, the ratio of the largest to the smallest approximating $10^6$, it is essential that some such degree of preciseness be held. But, since the resulting values of $L$ is physically significant to only two or three decimal places, it is not unlikely that a study of the number of decimal places to be held in the various stages of computation would reveal that a smaller number than actually used still would be satisfactory. Inasmuch, however, as 15-place tables were employed and the calculation was performed on a 10-row electric calculating machine, the added labor is negligible, and there is no uncertainty as to the effect of cumulative errors on the accuracy of the first several digits of $L$.

Example 2. The problem is to calculate the geometric mean distances and the associated inductances of a bus having cross section as in Fig. 2 with dimensions and currents equal, respectively, to the external dimensions ($a=6$ inches, $c=4.5$ inches and $e=3$ inches) and currents of the bus of example 1.

Substituting these values in Eqs. (9)–(11) and analogous equations and performing the indicated arithmetic yields

$$\log D_{xx} = 1.2504, \quad \log D_{xy} = 1.237,$$
$$\log D_{yy} = 0.9628, \quad \log D_{xz} = 1.232,$$
$$\log D_{zz} = 0.5573, \quad \log D_{yz} = 0.946.$$

Utilizing these values in Eq. (12) and analogous equations yields

$L_x = -0.0328+j0.002$ abhenry per centimeter of bus length,

$L_y = 0.2574-j0.502$ abhenry per centimeter of bus length,

$L_z = 1.0634+j0.50$ abhenry per centimeter of bus length.

Confirmatively, to three decimal places the same values of geometric mean distance and of associated inductance are yielded by the values of Table I conjoined with Eqs. (14)–(22).

Finally, from Eqs. (23)–(25) we have

$L_x = -0.030$ abhenry per centimeter of bus length,

$L_y = 0.257-j0.497$ abhenry per centimeter of bus length,

$L_z = 1.069+j0.498$ abhenry per centimeter of bus length.

Example 3. The problem is to calculate the reactive voltage drops on the conductors of the busses for examples 1 and 2.

The voltage drops can be calculated from

Eqs. (26), (27) and (28) by substituting therein the values of logarithmic mean distances calculated in examples 1 and 2. Alternatively, they can be calculated by multiplying the associated inductances by appropriate values of *jwI*. Utilizing the latter procedure and using inductances calculated from Eq. (7) yields

$E_x = 2jw(0.0141+j0.0209)$ abvolt per centimeter of bus length,
$E_y = 2jw(0.190+j0.243)$ abvolt per centimeter of bus length,
$E_z = 2jw(0.6795+j0.286)$ abvolt per centimeter of bus length.

Using inductances calculated from Eqs. (12) and analogous equations and from (20), (21) and (22) yields

$E_x = 2jw(0.0082+j0.0142)$ abvolt per centimeter of bus length,
$E_y = 2jw(0.153+j0.237)$ abvolt per centimeter of bus length,
$E_z = 2jw(0.532+j0.250)$ abvolt per centimeter of bus length.

Using inductances calculated from Eq. (23)–(25) yields

$E_x = 2jw(0.0075+j0.013)$ abvolt per centimeter of bus length,
$E_y = 2jw(0.151+j0.236)$ abvolt per centimeter of bus length,
$E_z = 2jw(0.532+j0.249)$ abvolt per centimeter of bus length.

Comparison of the values calculated from the approximate Eqs. (20)–(22) with those calculated from the simple approximate Eqs. (23)–(25) indicates that in general the simple approximate equations furnish sufficiently accurate values. Comparison of both sets with the values calculated from the general Eq. (7) reveals that the former are somewhat higher. Though these values would be sufficiently accurate for preliminary—and in most instances, final—design purposes, it is yet desirable to have means of calculating somewhat more accurate values without incurring the heavy labor incident to calculation with Eq. (7). Such a means of calculation is suggested by the writers' previous work[6] on single-phase coaxial square tubular busses. Therein it is shown that a bus comprised of coaxial circular tubular conductors has an inductance which differs little from that of a bus comprised of coaxial square tubular conductors having (1) the same thicknesses and (2) internal and external dimensions equal to the internal and external diameters of the corresponding circular conductors. Analogously, it is to be expected that the inductances of similarly dimensioned three-phase busses will be in close agreement. If so, it is to be conjectured that this agreement is better than that between corresponding values for the square tubular conductors calculated from the given geometry and from the approximate assumption of line-segment cross sections. That such is true is substantiated by the following illustrative examples.

Example 4. The problem is to calculate the inductances and reactive voltage drops of the conductors of a bus having cross section as in Fig. 3 with dimensions and currents identical with those of example 1.

Calculating the pertinent geometric mean distances as detailed in Appendix IV, substituting as indicated in Eqs. (1) and (3) and (26)–(28), and collecting terms yields the desired values:

$L_x = -0.023$ abhenry per centimeter of bus length,
$L_y = 0.271 - j0.537$ abhenry per centimeter of bus length,
$L_z = 1.384 + j0.537$ abhenry per centimeter of bus length,

and

$E_x = 2jw(0.006+j0.010)$ abvolt per centimeter of bus length,
$E_y = 2jw(0.165+j0.252)$ abvolt per centimeter of bus length,
$E_z = 2jw(0.692+j0.269)$ abvolt per centimeter of bus length.

Example 5. The problem is to calculate the inductances and reactive voltage drops of the conductors of a bus having cross section as in Fig. 4 with dimensions and currents identical with those of example 2.

Calculating the pertinent geometric mean distances as detailed in Appendix IV, substituting as indicated in Eqs. (1)–(3) and (26)–(28), and collecting terms yields the desired values:

$L_x=0$ abhenry per centimeter of bus length,
$L_y=0.297-j0.498$ abhenry per centimeter of bus length,
$L_z=1.099+j0.498$ abhenry per centimeter of bus length,

and

$E_x=0$ abvolt per centimeter of bus length,
$E_y=2jw(0.144+j0.249)$ abvolt per centimeter of bus length,
$E_z=2jw(0.549+j0.249)$ abvolt per centimeter of bus length.

Comparison of the two sets of values calculated in examples 4 and 5 with the corresponding first two sets of example 3 reveals

(i) That—corroborative of expectation—corresponding values for square and for circular conductors are in good agreement;

(ii) That—corroborative of conjecture—the values of example 4 afford a somewhat better approximation to the first set of values in example 3 than do the second set of values of example 3, calculated on the assumption of line-segment cross section.

### VII. CORRECTIONS FOR ROUNDED EDGES

To eliminate excessive edge heating caused by skin effect, square tubular conductors of commercial manufacture have rounded edges. Thus square tubular conductors of the sizes specified in examples 1 and 3 have rounded edges of outer, inner and mean radius indicated as follows:[16] $a=6$ inches; $t=(a-b)/2=0.2$ inch; $r_0=0.75$ inch; $r_i=0.55$ inch; $r_m=0.65$ inch, $c=4.5$ inches; $t'=(c-d)/2=0.25$ inch; $r_0=0.625$ inch; $r_i=0.375$ inch; $r_m=0.5$ inch, $e=3$ inches; $t''=(e-f)/2=0.9$ inch; $r_0=1$ inch; $r_i=0.1$ inch; $r_m=0.55$ inch.

Dwight and Wang[10] have derived, semi-empirically, that the approximate increase in self geometric mean distance caused by rounding the edges of a conductor of square cross section is

$$\Delta \log D_s=-[k^2\log k^{-1}+0.716k^2+0.5k^3+\cdots] \text{ abhenrys per centimeter of conductor length,} \tag{32}$$

where $k=0.463r_m/s$, $r_m$ is the mean radius of the rounded edge, $s$ the mean side of the square cross section, and $r_m$ is small compared with $s$, the accuracy increasing as $r_m/s\rightarrow 0$.

From Eqs. (1)–(3) it follows that the approximate increase in associated inductance is

$$\Delta L=2[k^2\log k^{-1}+0.716k^2+0.5k^3+\cdots] \text{ abhenrys per centimeter of conductor length.} \tag{33}$$

For conductor $x$, $s=a-t=5.8$ inches and $r_m=0.65$ inch. Hence, $k=0.052$, $\Delta L_x=0.020$, and $\Delta E_x(=jwI_x\Delta L_x)=2jw(-0.0049-j0.0093)$.

For conductor $y$, $s=c-t'=4.25$ inches and $r_m=0.5$ inch. Hence, $k=0.0545$, $\Delta L_y=0.0216$, and $\Delta E_y=-2jw(-0.005+j0.009)$.

For conductor $z$, $s=e-t''=2.1$ inches and $r_m=0.55$ inch. Hence, $k=0.122$, $\Delta L_z=0.011$, and $\Delta E_z=2jw(0.006)$.

Adding these corrections to the first set of values of examples 1 and 3 yields

$L_x=-0.029$ abhenry per centimeter of bus length,
$L_y=0.252-j0.572$ abhenry per centimeter of bus length,
$L_z=1.370+j0.572$ abhenry per centimeter of bus length,

and

$E_x=2jw(0.009+j0.012)$ abvolt per centimeter of bus length,
$E_y=2jw(0.185+j0.252)$ abvolt per centimeter of bus length,
$E_z=2jw(0.686+j0.286)$ abvolt per centimeter of bus length.

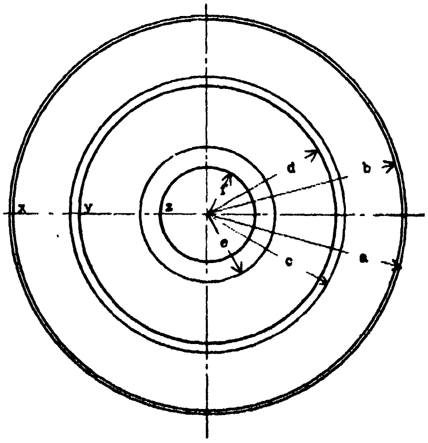


FIG. 3. Cross section of coaxial bus comprised of circular tubular conductors.

---

[16] *Chase Electrical Handbook* (Chase Brass and Copper Company, Inc., Waterbury, Connecticut, 1941), pp. 80–81.

The facts that rounding the corners of a conductor will result in a change of associated inductance and that the changes just calculated from Eq. (33) are of the correct order of magnitude are confirmed by comparing, in turn, the first set of values of example 1 (for the bus of Fig. 1), the values just listed (for a bus stemming from that of Fig. 1 through rounding the corners of the conductors), and the values of example 4 (for the bus of Fig. 3). Now a circular tubular conductor can be considered the limiting case of a rounded square tubular conductor, its internal and external sides being equal to the internal and external diameters of the circular conductor. Hence the values just listed should be—as they are—intermediate between the corresponding values for the square conductors of Fig. 1 and the circular conductors of Fig. 4.

## VIII. CORRECTIONS FOR SKIN AND PROXIMITY EFFECTS

As the frequency increases from zero to a specified value the currents in the conductors of the bus with cross section as in Fig. 3 tend to concentrate in thin surface strata on the adjacent surfaces of the conductors: the inner surface of conductor $x$ and the outer surface of conductor $y$, the inner surface of conductor $y$ and the outer surface of conductor $z$. This phenomenon results in a change of the associated inductance (and resistance) of each conductor. Detailed mathematical analysis[6, 11, 17] reveals that at 60-cycles the changes in the associated inductances of practically dimensioned busses such as those of example 4 are negligible.

In view of these negligible changes in the inductances of this coaxial circular tubular bus, the small differences between the inductances of this bus and those of the coaxial square tubular bus of example 1—especially, after applying corrections for the rounded edges, and the general geometric similarity between these two busses, we conclude that in practically dimensioned rounded square tubular busses in general 60-cycle skin and proximity effects are negligible insofar as the inductances of the bus are concerned. This conclusion is in accord with the results of analytic and experimental investigations of the inductances of single-phase coaxial[6, 11] and coplanar busses comprised of square tubular or rectangular conductors.

On the other hand, the reactive voltage drops will be somewhat, though slightly, changed by skin and proximity effects. This stems from the fact that the resistances of the conductors are changed by skin and proximity effects. Therefore the current distribution in the conductors, and thus the resistive and reactive voltage drops in the conductors, changes with frequency. At 60-cycles and for bus and conductor geometry commonly encountered in practice it commonly suffices to calculate the a.c. resistances of the square tubular conductors by multiplying their d.c. resistance $R_{\text{d.c.}}$ by the known[7, 8, 16] skin-effect factor $K_s$.

Accurate calculation of $R_{\text{a.c.}}$ requires that the values of $K_sR_{\text{d.c.}}$ thus obtained be multiplied, in turn, by appropriate proximity effect factors $K_p$. Because of the intractable mathematics involved, it has not been possible to calculate rigorously these factors for single-phase or three-phase coaxial square tubular conductors. On the other hand, the simpler—though yet mathematically difficult—corresponding problem for coaxial circular tubular conductors has been solved. In view of the above discussion and analysis of the corresponding inductances of these two busses, and in view of their general geometric similarity, it is to be expected that values of $K_p$ for a bus comprised of circular tubular conductors would be substantially the same as values of $K_p$ for a bus comprised of square tubular conductors having sides and thicknesses identical with the diameters and thicknesses of the corresponding circular conductors.

Now Dwight[17] has published equations ((22), (23), and (29) of Reference 17) for the ratio of $R_{\text{a.c.}}/R_{\text{d.c.}}(\equiv K_sK_p)$ for each of the three conductors of a three-phase coaxial circular tubular bus and curves (Fig. 2 of Reference 17) for the particular case of balanced currents. Accordingly, if a value obtained from Dwight's equations or curves is divided by $K_s$ for the same *circular tubular conductor*, there is obtained a value of $K_p$ which is a good approximation to the $K_p$ of the corresponding square tubular conductor. (It is to be noted that the required value

---

[17] H. B. Dwight, "Reactance and skin effect of concentric tubular conductors," Trans. A.I.E.E. 61, 513–518, 976–977 (1942).

of $K_s$ can be obtained from Dwight's equations or curves: it is equal to the value of $K_sK_p \equiv K_s$ for the conductor when considered as the innermost conductor of a three-phase bus.) Commonly, however, the refinement stemming from use of the values of $K_p$ is not necessary in practice.

### IX. BUSSES COMPRISED OF STRUCTURAL TUBULAR CONDUCTOR

A conductor comprised of two slightly separated channels or angles, placed flange to flange, usually is not of a square cross section. If, however, the cross section does not differ too greatly from a square cross section, and commonly this is the case in practice, it is to be expected that a fairly good approximation to the actual inductance of the bus can be effected by judicious use of the equations for the square cross section. In an earlier paper[6] on single-phase coaxial busses the writers calculated that an equivalent square tubular bus with conductors having either the same external perimeters (rounded edges aside) or sides equal to the webs of the given structural conductors have inductances in good agreement with those of the structural tubular bus and showed that these analytic deductions were confirmed by experiment. It is to be inferred that selecting an equivalent three-phase bus on one or the other of the bases mentioned will yield approximate values of inductance within a few percent of the actual values, which is all that is required for most design purposes.

Skin-effect and proximity-effect factors for calculating the a.c. resistances of the conductors can be obtained in like manner.

### X. BALANCING CONDUCTOR CURRENTS IN THREE-PHASE COAXIAL BUSSES

Because of the unbalanced geometry of the three-phase coaxial bus the phase currents will, in general, not be balanced. In such case they must be calculated by the procedure detailed in chapter 2 of reference 13. For obvious reasons, however, balance is most desirable. Though a considerable degree of balance can be obtained by proper geometrical design (make, in so far as is feasible, the associated impedances identical), complete balance is not to be so obtained. The system can be balanced, however, by the use of suitably designed and suitably placed metal sleeves fitted around the conductors. Though account of this scheme has long been published,[6, 18] and satisfactory results have been obtained in practice,[19] it would appear that comparatively little use of it has been made to date.

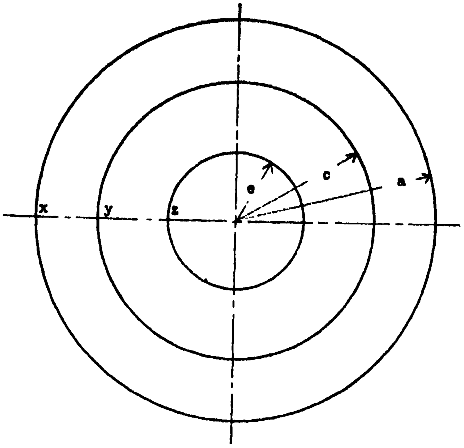


FIG. 4. Cross section of coaxial bus comprised of circular tubular line-segment conductors.

### APPENDIX I

The cross section $S_x \equiv S_1 - S_2$ of conductor $x$ of Fig. 1 can be considered as the symmetrical superposition of a positive square area $S_1$ of side $a$ and a negative square area $-S_1$ of side $b$. By well known theory[6, 12]

$$S_x^2 \log D_{xx} = S_1^2 \log D_{11} + S_2^2 \log D_{22} - 2S_1S_2 \log D_{12} \quad (34)$$

wherein as detailed in Appendix I of reference 6

$$S_1^2 \log D_{11} = -(25/12)S_1^2 - (1/6)[K(a, a) + K(0, 0) - 2K(a, 0)], \quad (35)$$

$$S_1S_2 \log D_{12} = -(25/12)S_1S_2 - (1/6)\{K[(a+b)/2, (a+b)/2] + K[(a-b)/2, (a-b)/2] - 2K[(a+b)/2, (a-b)/2]\} \quad (36)$$

and, generally,

$$K(A, B) = (A^4 - 6A^2B^2 + B^4) \log(A^2+B^2)^{\frac{1}{2}} - 4AB^3 \tan^{-1}(A/B) - 4A^3B \tan^{-1}B/A. \quad (37)$$

Substituting appropriately in Eq. 34, col-

---

[18] V. G. Egorov, "Equalizing currents in separate lines in complex bus bars," Elektrichestyo, 244–245 (1930).

[19] W. K. Boice, "Current-balancing reactors for turbine generators," Power Plant Eng. 48, October, 94, 120 (1944).

lecting corresponding terms, and solving for $\log D_{xx}$ yields Eq. (4).

In similar fashion, the cross section $S_y \equiv S_3 - S_4$ of conductor $y$ of Fig. 1 can be considered as the symmetrical superposition of a positive square area $S_3$ of side $c$ and a negative square area $-S_4$ of side $d$. By well-known theory

$$S_xS_y \log D_{xy} = S_1S_3 \log D_{13} - S_1S_4 \log D_{14} - S_2S_3 \log D_{23} + S_2S_4 \log D_{24}, \quad (38)$$

wherein the four terms on the right-hand side of Eq. 38 are analogous to those of Eq. (36).

Substituting appropriately in Eq. (38), collecting corresponding terms, and solving for $\log D_{xy}$ yields Eq. (5).

## APPENDIX II

The cross section $S_x$ of Fig. 2 is comprised of four identical line segments, $S_1$, $S_2$, $S_3$ and $S_4$. By well-known theory

$$S_x^2 \log D_{xx} = 4(S_{11}^2 \log D_{11} + 2S_1S_2 \log D_{12} + S_1S_3 \log D_{13}), \quad (39)$$

wherein as detailed in Appendix III of reference 6,

$$\log D_{11} = \log a - 3/2, \quad (40)$$

$$\log D_{12} = \log a + (1/2) \log 2 + \pi/4 - 3/2, \quad (41)$$

$$\log D_{13} = \log a + \pi/2 - 3/2. \quad (42)$$

Substituting appropriately in Eq. (39), collecting corresponding terms, and solving for $\log D_{xx}$ yields Eq. (9).

In similar fashion, the cross section $S_y$ of Fig. 2 is comprised of four identical line segments $S_5$, $S_6$, $S_7$, and $S_8$. By well known theory

$$S_xS_y \log D_{xy} = 4(S_1S_5 \log D_{15} + 2S_1S_6 \log D_{16} + S_1S_7 \log D_{17}), \quad (43)$$

wherein as detailed in appendix III of reference 6,

$$\log D_{15} = (1/2)\{\log(a^2+c^2) + [(a^2-c^2)/ac] \times \tan^{-1}[(a+c)/(a-c)] - \pi(a-c)^2/4ac - \log 2 - 3\}, \quad (44)$$

$$\log D_{16} = (1/4)\{[(a+c)^2/ac] \log(a+c) - [(a-c)^2/ac] \log(a-c) - 2 \log 2 + \pi - 6\}, \quad (45)$$

$$\log D_{17} = (1/2)\{\log(a^2+c^2) - [(a^2-c^2)/ac] \tan^{-1}[(a-c)/(a+c) + \pi(a+c)^2/4ac - \log 2 - 3\}. \quad (46)$$

Substituting appropriately in Eq. 43, collecting corresponding terms and solving for $\log D_{xy}$ yields Eq. 10.

## APPENDIX III

The values of $K(r)$ in Table I are obtained by substituting the indicated values of $A$ and $r$ in Eq. (13), obtaining the values of the logarithmic and arctangentic terms from the Work Projects Administration Tables, and performing all computation on a 10-row electric calculating machine. To avoid cumulative errors all computation was effected to eight significant figures. For actual use, however, $K(r)$ need only to be known to the three or four significant figures given in Table I. For use in design, the designer may find it convenient to plot a curve of $K(r)$ *versus* $r$ from the values given.

TABLE I.

| $A$ | $r$ | $K(r)$ | $A$ | $r$ | $K(r)$ |
|---|---|---|---|---|---|
| 0 | 1 | −0.541 | 0.50 | 1/3 | −0.5611 |
| 0.10 | 9/11 | −0.553 | 0.55 | 7/31 | −0.5611 |
| 0.15 | 17/23 | −0.5560 | 0.60 | 1/4 | −0.5612 |
| 0.20 | 2/3 | −0.558 | 0.65 | 5/33 | −0.5612 |
| 0.25 | 3/5 | −0.5590 | 0.70 | 3/17 | −0.5612 |
| 0.30 | 7/13 | −0.5597 | 0.75 | 1/7 | −0.5612 |
| 0.35 | 13/27 | −0.5603 | 0.80 | 1/9 | −0.5612 |
| 0.40 | 3/7 | −0.5607 | 0.90 | 1/19 | −0.5612 |
| 0.45 | 11/29 | −0.5608 | 1.0 | 0 | −0.5612 |

## APPENDIX IV

The self geometric mean distance of a circular annulus is[18]

$$\log D_s = \log r_1 - [4(r_1^2 - r_2^2)^2]^{-1}[r_1^4 - 4r_1^2r_2^2 + r_2^4(3 + 4 \log r_1/r_2)], \quad (47)$$

wherein $r_1$ and $r_2$ are, respectively, the external and internal radii of the annulus.

The mutual geometric mean distance of a circular annulus and an area of arbitrary area within it is

$$\log D_m = (r_1^2 - r_2^2)^{-1}(r_1^2 \log r_1 - r_2^2 \log r_2) - 1/2. \quad (48)$$

Substituting appropriately in Eqs. (47) and (48) yields the self and mutual geometric mean distances of Fig. 3:

$$\log D_{xx} = 1.0764, \quad \log D_{xy} = 1.0649,$$
$$\log D_{yy} = 0.7739, \quad \log D_{xs} = 1.0649,$$
$$\log D_{ss} = 0.2174, \quad \log D_{ys} = 0.7543.$$

Substituting these values and those of the currents in Eqs. (1)–(3) and (26)–(28) yields the values of example 4. Confirmatively the values of reactive voltage drop on conductors $y$ and $z$ are in agreement with the values obtained by Dwight[17] who utilized the same bus to illustrate his paper. The value for $I_x$ is considerably smaller than Dwight's value. As, however, this small value stems from the algebraic summation of several large values, the difference in the two values is not surprising if (as the writers understand) Dwight's values were obtained by slide rule whereas the writers' were obtained by use of an electric calculating machine, all numerical values being held to eight decimal places during computation.

The self-geometric mean distance of a circle and the mutual geometric mean distance of the circle to any arbitrary area within it are identical:

$$\log D_s = \log D_m = \log r_1,$$

where $r_1$ is the radius of the circle.

Substituting appropriately in Eq. (49) yields the self- and mutual geometric mean distances of Fig. 4:

$$\begin{aligned} \log D_{xx} &= 1.0986, & \log D_{xy} &= 1.0986, \\ \log D_{yy} &= 0.8109, & \log D_{xz} &= 1.0986, \\ \log D_{zz} &= 0.4055, & \log D_{yz} &= 0.8109. \end{aligned}$$

Substituting these values and those of the currents in Eqs. (1)–(3) and (26)–(28) yields the values of example 5.

## Calendar of Meetings

**November, 1947**

| | |
|---|---|
| 3–5 | National Electronics Conference, Chicago, Illinois |
| 3–7 | American Institute of Electrical Engineers, Chicago, Illinois |
| 6–7 | Society of Automotive Engineers (Fuels and Lubricants Meeting), Tulsa, Oklahoma |
| 7–8 | Conference on X-Ray and Electron Diffraction, Mellon Institute of Industrial Research, Pittsburgh, Pennsylvania |
| 9–12 | American Institute of Chemical Engineers, Detroit, Michigan |
| 17–19 | Institute of Radio Engineers and the Radio Manufacturers Association, Engineering Department, Rochester, New York |
| 28–29 | American Physical Society, Houston, Texas |

**December, 1947**

| | |
|---|---|
| 1–3 | Society of Automotive Engineers (Air Transport Engineering Meeting), Kansas City, Missouri |
| 1–5 | American Society of Mechanical Engineers, New York, New York |
| 26–31 | American Association for the Advancement of Science, Chicago, Illinois |
| 29–31 | American Physical Society, Chicago, Illinois |

**January, 1948**

| | |
|---|---|
| 2–3 | American Physical Society, Los Angeles, California |
| 21–23 | American Society of Civil Engineers, New York, New York |

# Mechanical Behavior of High Damping Metals*

CLARENCE ZENER
*Institute for the Study of Metals, The University of Chicago, Chicago, Illinois*

(Received July 16, 1947)

The relation between the various measures of internal friction are independent of the precise mechanism of the dissipation of energy when the internal friction is small, but not when it is large. In this paper the relation between the two measures most commonly used, logarithmic decrement and tangent of the angle with which strain lags behind stress, is deduced for all levels of internal friction in the important case in which the dissipation of energy is due to a relaxation process having a single time of relaxation. The conditions are further derived under which a specimen of such a metal will not vibrate, but returns aperiodically to its equilibrium configuration.

## I. INTRODUCTION

SEVERAL measures of internal friction are in common use. Engineers frequently report the specific damping capacity, $\Delta E/E$, defined as the ratio of the energy dissipated per cycle over the average energy of vibration. The measure generally used in physics varies according to the method used in its determination. When the method of free decay is employed, the logarithmic decrement $\delta$ is used. On the other hand, when the method of forced oscillation is employed, the angle $\alpha$ by which strain lags behind stress is usually measured, and the internal friction is reported in terms of $\tan\alpha$. At low values of internal friction all these various measures are related by simple conversion factors. Thus[1]

$$\Delta E/E = 2\delta = 2\pi \tan\alpha. \qquad (1)$$

Recent measurements[2] in this laboratory have indicated that polycrystalline metals have, under appropriate conditions, a $\tan\alpha$ as large as 0.1. Under these conditions the term specific damping capacity becomes inappropriate since the very concept of energy of vibration is ambiguous. Under these conditions the relation between $\delta$ and $\tan\alpha$ in Eq. (1) begins to become inaccurate size in its derivation $\tan\alpha$ was assumed to be a small order quantity. The precise relation between $\delta$ and $\tan\alpha$ is dependent upon the mechanism of the dissipation of energy. One important mechanism of energy dissipation in metals arises from the relaxation of an internal parameter which is disturbed from its equilibrium value by a change in strain.[3-6] This parameter could refer, for example, to the stress induced difference in chemical potential of solute atoms in two types of interstitial positions. Frequently this internal parameter, which will be denoted by $A$, obeys a simple relaxation equation of the type

$$dA/dt = -\tau_\epsilon^{-1}A - \gamma\dot{x}, \qquad (2)$$

where $x$ denotes the deformation and $\tau_\epsilon$ denotes the time of relaxation at constant strain. In such cases the relation between stress and strain is completely specified, and hence one may deduce, the precise relation between $\delta$ and $\tan\alpha$ obtained from free-oscillation and forced-oscillation experiments, respectively. This relation is obtained in section II. In section III the conditions are derived under which the material behaves in an aperiodic manner, that is, under which the material is self-overdamped.

## II. RELATION BETWEEN $\delta$ AND TAN$\alpha$

In the most appropriate method of investigating highly damping material one employs an auxiliary-inertia member so that the inertia of the specimen itself is of no consequence. One thereby avoids all uncertainty as to the precise manner in which strain is distributed throughout the specimen. The strain is everywhere specified

---

* This research has been supported by O. N. R., Contract No. N6ori-20-IV.

[1] C. Zener, Phys. Rev. 52, 230 (1937).

[2] T'ing-Sui Kê, Phys. Rev. 71, 533 (1947).

[3] C. Zener, W. Otis and R Nuckolls, Phys. Rev. 53, 100 (1938).

[4] R. H. Randall, F. C. Rose, and C. Zener, Phys. Rev. 56, 343 (1939).

[5] J. Snoek, Physica 8, 711 (1941); 9, 862 (1942).

[6] C. Zener, Phys. Rev. 71, 34 (1947).

by a single-deformation parameter $x$ which describes the overall deformation, e.g., angle of twist in a torsion experiment. We shall denote by $f$ the force conjugate to $x$, e.g., the torque acting upon specimen in a torsion experiment.

The deformation $x$ is a linear function of $f$ and the internal parameter $A$. Thus

$$x = M_R^{-1}f + \lambda A. \tag{3}$$

Here $M_R$ refers to the elastic modulus under conditions of complete relaxation, and is called the relaxed elastic modulus. We shall find it convenient to also introduce the unrelaxed modulus $M_U$, which refers to the elastic modulus measured under conditions in which no relaxation takes place. Upon using Eqs. (2) and (3) we see that

$$M_U/M_R = 1 + \gamma\lambda. \tag{4}$$

We shall also find it convenient to introduce the time of relaxation of $A$ under conditions of constant stress, $\tau_\sigma$. Again comparing Eqs. (2) and (3) we find

$$\tau_\sigma/\tau_\epsilon = 1 + \gamma\lambda. \tag{5}$$

Making use of these definitions we find that elimination of $A$ between Eqs. (2) and (3) leads to

$$dx/dt + \tau_\sigma^{-1}x = M_U^{-1}(df/dt + \tau_\epsilon^{-1}f). \tag{6}$$

The angle $\alpha$ by which strain lags behind stress in a forced oscillation experiment may now be found by substituting

$$f = f_0 e^{i\omega t}$$

into Eq. (6). We find

$$\tan\alpha = \frac{M_U - M_R}{(M_U \cdot M_R)^{\frac{1}{2}}} \cdot \frac{\omega(\tau_\epsilon\tau_\sigma)^{\frac{1}{2}}}{1 + \omega^2 \cdot \tau_\epsilon\tau_\sigma}. \tag{7}$$

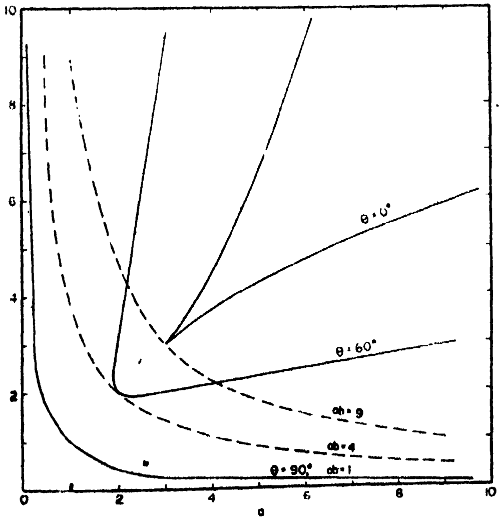


**Fig. 1.** Diagram for determination of conditions for $\delta_m$. Logarithmic decrement is constant along lines of constant $\theta$. Variation of frequency of vibration corresponds to motion along a line of constant $ab$.

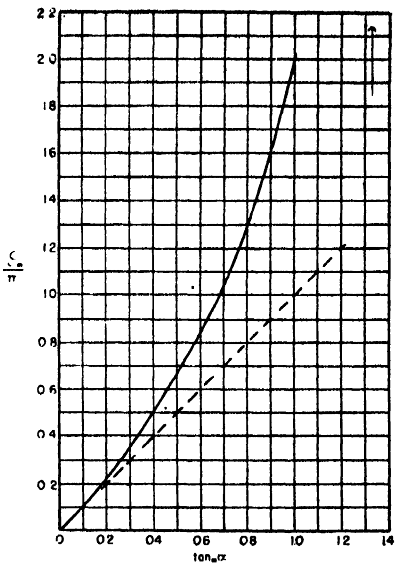


**Fig. 2.** Relation between maxima values of $\delta$ and of $\tan\alpha$ in a plot *vs.* frequency of vibration. Solid line gives exact relation. Dashed line gives extrapolation of relation valid at small values of internal friction.

In the case of a free oscillation experiment, the force $f$ in Eq. (6) is the negative of the force which acts upon the inertia member. Thus if we denote by $I$ the inertia,

$$Id^2x/dt^2 = -f. \tag{8}$$

An equation containing $x$ alone is now obtained by eliminating $f$ between Eqs. (6) and (8). We thus find

$$d^3x/dt^3 + \tau_\epsilon^{-1}d^2x/dt^2 + \omega_U^2(dx/dt + \tau_\sigma^{-1}x) = 0. \tag{9}$$

Here we have introduced the "unrelaxed" angular frequency

$$\omega_U = (M_U/I)^{\frac{1}{2}}. \tag{10}$$

We shall later find it convenient to introduce the "relaxed" angular frequency

$$\omega_R = (M_R/I)^{\frac{1}{2}}. \tag{11}$$

The logarithmic decrement is now found by sub-

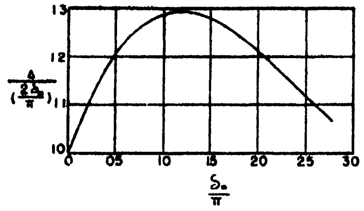


FIG. 3. Relation between relaxation strength and $\delta_m/\pi$.

stituting

$$x=x_0e^{-\xi t} \tag{12}$$

into Eq. (9). According as $\xi$ is expressed as

$$\xi=\xi_0e^{i\theta} \tag{13}$$

or as

$$\xi=\mu+i\eta \tag{14}$$

the logarithmic decrement is given by

$$\delta/\pi=2\cot\theta \tag{15}$$

or

$$\delta/\pi=2\mu/\eta. \tag{16}$$

Upon using the substitutions

$$E^3=\omega_U{}^2\tau_\sigma{}^{-1},\quad a=(\tau_\epsilon E)^{-1},\quad b=\omega_U{}^2E^{-2},\quad U=E^{-1}\xi,$$

we transform this cubic equation into the standard form studied by Jahnke and Emde,[7] namely

$$U^3-aU^2+bU-1=0. \tag{17}$$

As a clue as how to proceed to evaluate the logarithmic decrement from this cubic equation we note that

$$ab=M_U/M_R=1+\gamma\lambda, \tag{18}$$

and

$$b/a=(\omega_U\omega_R\tau_\epsilon\tau_\sigma)^{\frac{1}{2}}. \tag{19}$$

In experiments upon a given material at a given temperature, the product $ab$ therefore remains fixed. On the other hand the ratio $b/a$ varies as we vary the frequencies $\omega_U$, $\omega_R$, either by changes in dimensions of the specimen or through changes in the inertia $I$. Upon comparing the manner with which the angle $\theta$, and hence $\delta$, varies in $(a, b)$ space as shown in Fig. 1 taken from Jahnke and Emde,[7] with how the product $ab$ varies, we see that for a given product $ab$ the logarithmic decrement is a maximum when the ratio $b/a$ is unity. Our problem is therefore considerably simplified if we restrict ourselves to finding the maximum value $\delta_m$ of the logarithmic decrement in a plot of $\delta$ *vs.* frequency of oscillation. In practice this maximum value may also be obtained by a variation of temperature over a range so small that the ratio $M_U/M_R$ remains essentially constant, but sufficiently large so that $\tau_\sigma$ and $\tau_\epsilon$ vary by a large factor. This restriction allows us to replace $b$ by $a$ in Eq. (17). We now note that this simplified cubic equation may be factorized as follows,

$$(U-1)\{U^2+(1-a)U+1\}=0. \tag{20}$$

The first root of Eq. (20) corresponds to an aperiodic motion. Upon solving for the two roots in the second factor, and upon noting the relations in Eqs. (14) and (16), we obtain

$$\delta_m/\pi=2(a-1)\{4-(a-1)^2\}^{-\frac{1}{2}}. \tag{21}$$

We shall now compare $\delta_m$ with the maximum value of $\tan\alpha$, $\tan_m\alpha$, obtained in a plot of $\tan\alpha$ *vs.* $\omega$. Upon observing that the parameter $a$ in Eq. (21) is given by

$$a^2=M_U/M_R, \tag{22}$$

we obtain from Eq. (7)

$$\tan_m\alpha=(1/2)(a-a^{-1}). \tag{23}$$

A plot may now be made of $\delta_m/\pi$ *vs.* $\tan_m\alpha$ through the two parametric equations (21) and (23). Such a plot is presented as Fig. 2. It is to be observed that $\delta_m$ becomes infinite when $a$ is 3, and hence when $\tan_m\alpha$ is 4/3, or when $\alpha$ is 53°. An analytic relation may be found between $\delta_m/\pi$ and $\tan_m\alpha$ when the internal friction is

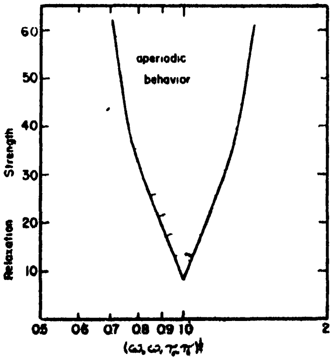


FIG. 4. Conditions for aperiodic behavior.

---

[7] E. Jahnke and F. Emde, *Funktionentafeln* (Teubner, Leipzig, 1938), second edition, addenda, pp. 20–23.

small. Towards this end we introduce the quantity

$$y = a - 1,$$

and expand $\delta_m/\pi$ and $\tan_m\alpha$ in a power series in $y$. We obtain

$$\delta_m/\pi = y(1 + y^2/8 + \quad).$$
$$\tan_m\alpha = y(1 - y/2 + \quad)$$

and hence

$$\tan_m\alpha = (\delta_m/\pi)\{1 - \delta_m/2\pi + \quad\}.$$

The quantity

$$\Delta = (M_U - M_R)/M_R$$

is known as the relaxation strength, and is of theoretical significance since it is directly related to the constants of the material. Thus from Eq. (4)

$$\Delta = \gamma\lambda.$$

In terms of the parameter $a$ the relaxation strength may be written as

$$\Delta = a^2 - 1. \tag{24}$$

The parametric Eqs. (21) and (24) may now be used to obtain a plot of $\Delta$ *vs.* $\delta_m/\pi$. Such a plot is presented as Fig. 3. The analytic relation at small values of $\Delta$ is readily seen to be

$$\Delta = 2(\delta_m/\pi)\{1 + \delta_m/2\pi + \quad\}.$$

### III. CONDITIONS FOR APERIODIC BEHAVIOR

A specimen will not vibrate freely if all three roots of Eq. (17) are real. The boundary of the region in $(a,b)$ space for which all three roots are real may be found by the method outlined by Jahnke and Emde.[7] This boundary is labeled $\theta = 0$ in Fig. 1. A more useful plot is obtained when this boundary is expressed in terms of the two variables

$$\Delta = ab - 1$$

and

$$(\omega_U\omega_R\tau_\epsilon\tau_\sigma)^2 = (b/a)^{\frac{1}{2}}.$$

Such a plot has been deduced by graphical methods from an enlarged plot of the aperiodic boundary in $(a,b)$ space, and is presented as Fig. 4.

---

# The Field of a Microwave Dipole Antenna in the Vicinity of the Horizon. II*

C. L. Pekeris**

*Columbia University Mathematical Physics Group, Columbia University, New York, New York*

(Received July 20, 1947)

In this paper the results of a previous investigation are extended to include cases where the elevation of the transmitter $z_1$, or of the receiver $z$, is less than about 1 when expressed in natural units of height. When $z_1 < 1$ and $z > 1$ the electromagnetic potential $\Psi$ in the vicinity of the horizon is given, in the case of strong absorption $(\beta \gg 1)$, by

$$|\Psi| = [(\tfrac{3}{2})^{\frac{1}{2}}/\pi(r\bar{r})^{\frac{1}{2}}]\,|(1+\beta z_1)G(p)/\beta|, \tag{a}$$

where

$$\beta_v = i(k_0/q)^{\frac{1}{3}}(\epsilon_1 - 1)^{\frac{1}{2}}/\epsilon_1;\quad \beta_h = \beta_v\epsilon_1, \tag{b}$$

and the notation of $I$ is used. Expression (a) equals expression $(B)$ or $(C)$ of $I$ multiplied by the factor $|(1+\beta z_1)|$.

When both $z_1$ and $z$ are less than about unity the potential is given by

$$|\Psi| = (L/r^2)\,|(1+\beta z_1)(1+\beta z)g(p')/\beta^2|, \tag{c}$$

where $L$ is the natural unit of horizontal distance. Expression (c) is equal to the expression for the surface wave given in $(D)$ of $I$ multiplied by the factor $|(1+\beta z_1)(1+\beta z)|$.

In I[1] closed expressions were obtained for the field in the vicinity of the horizon in the three cases: (1) $z_1 > 1$, $z > 1$; (2) $z_1 = 0$, $z > 1$; and (3) $z_1 = 0$, $z = 0$. The purpose of this investigation is to extend the analysis to cover the cases (4) $z_1 < 1$, $z > 1$ and (5) $z_1 < 1$, $z < 1$. It was shown in Eqs. (16) and (17) of I that the Hertzian potential $\Psi$ of a dipole situated at elevation $z_1$ in an atmosphere of standard refraction is given by

$$\Psi_1 = (-i\pi/3)e^{i\omega t}(qk_0^2)^{-\frac{1}{3}}\int_0^\infty J_0(kr)kdkN(u)$$
$$\times[M(u_1) + (B/A)N(u_1)],\quad z > z_1, \tag{1}$$

---

* This paper is based on work done for the Naval Research Laboratory under contract N6ori-110, task order No. 3, with the Navy's Office of Research and Inventions. Publication assisted by the Ernest Kempton Adams Fund for Physical Research of Columbia University.

** On leave of absence from Massachusetts Institute of Technology.

[1] C. L. Pekeris, J. App. Phys. 18, July, 1947. This paper will be referred to as I.

$$\Psi_2=(-i\pi/3)e^{i\omega t}(qk_0^2)^{-\frac{1}{3}}\int_0^\infty J_0(kr)kdkN(u_1)$$
$$\times[M(u)+(B/A)N(u)],\quad 0<z<z_1,\quad (2)$$

where

$$u(z)=u_0+z;\ u_0=(k_0^2-k^2)/H^2;\ k_0=2\pi/\lambda,\quad (3)$$
$$M(u)=u^{\frac{1}{2}}H_{1/3}^{(1)}[(2/3)u^{\frac{3}{2}}],$$
$$N(u)=u^{\frac{1}{2}}H_{1/3}^{(2)}[(2/3)u^{\frac{3}{2}}],\quad (4)$$
$$B/A=-[\beta M(u_0)-\dot{M}(u_0)]/Q(u_0);$$
$$Q(u_0)=\beta N(u_0)-\dot{N}(u_0).\quad (5)$$

In order to evaluate (1) for small values of $z_1$ we expand the integrand into a power series in $z_1$ as follows:

$$\begin{aligned}Q(u_0)&[M(u_1)+(B/A)N(u_1)]\\ &=\beta[M(u_1)N(u_0)-N(u_1)M(u_0)]\\ &\quad-[M(u_1)\dot{N}(u_0)-N(u_1)\dot{M}(u_0)]\\ &=(6i/\pi)[(1+\beta z_1)-u_0z_1^2(\beta z_1+3)/6\\ &\qquad-z_1^3(\beta z_1+2)/12+\cdots],\end{aligned}\quad (6)$$

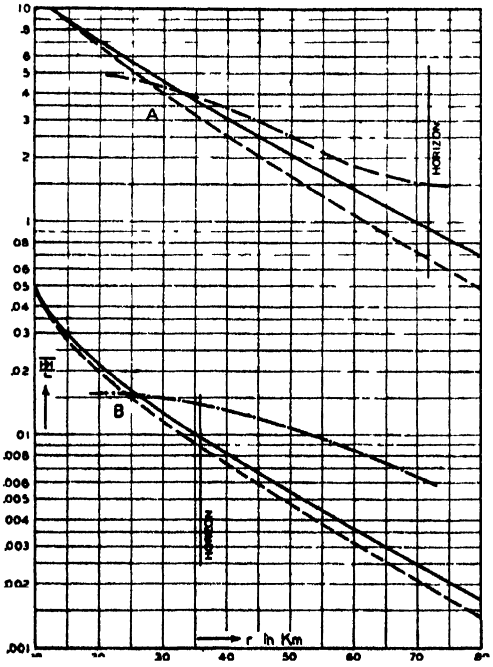


FIG. 1. Comparison of our results with those of B. van d. Pol and H. Bremmer.[2] $\lambda=700$ cm; A: $z_1=0.63$, $z=0.63$; B: $z_1=0$, $z=0.63$.

——— Exact values from normal-mode theory.
— — — — Eq. (20).
— · — · — Eqs. (A) and (B) of reference 1, valid when $z$ is large.

where use was made of the relations

$$N(x)\dot{M}(x)-M(x)\dot{N}(x)=(6i/\pi),$$
$$\frac{d^2N}{dx^2}+xN=0,\quad \frac{d^2M}{dx^2}+xM=0.\quad (7)$$

Since for the field in the vicinity of the horizon the principal contribution to the integrals in (1) and (2) arises from a small range of integration in the neighborhood of $k=k_0$, we may consider $\beta(=iv_0^{\frac{1}{3}}\epsilon_1^{-\frac{1}{2}})$ in (6) as a constant (to within $\beta_h^{-2}<10^{-5}$). Hence, on disregarding powers of $z_1$ above the first, we have

$$\Psi_1(z_1,z)\simeq(1+\beta z_1)\Psi(0,z),\quad (8)$$
$$\Psi(0,z)=2e^{i\omega t}(qk_0^2)^{-\frac{1}{3}}$$
$$\times\int_0^\infty J_0(kr)kdkN(u)/Q(u_0).\quad (9)$$

Expression (9), which is exact, has been evaluated in I for the case of strong absorption ($\beta\gg1$) and in the neighborhood of the horizon:

$$\Psi(0,z)\simeq[(3/2)^{\frac{1}{2}}/\pi(r\bar{r})^{\frac{1}{2}}]|G(p)/\beta|,$$
$$p=(3/2)^{\frac{1}{3}}(r-\bar{r})/L.\quad (10)$$

It can be shown that if the other terms in (6) are included, $G(p)$ in (10) must be replaced by

$$G(p)+i(3/2)^{\frac{1}{3}}z_1^2[(\beta z_1+3)/6(1+\beta z_1)]\frac{dG}{dp}$$
$$-z_1^3[(\beta z_1+2)/(1+\beta z_1)12]G(p)+\cdots.\quad (11)$$

Relation (8) agrees with a similar result arrived at by v. d. Pol and Bremmer[2] on the basis of the normal mode theory. Burrows and Gray[3] have also computed a function

$$F(L)=(3^{\frac{1}{2}}/2\pi)|G[(8/9)^{\frac{1}{3}}p]|,$$

using the normal-mode theory.

Turning now to the case where both $z_1$ and $z$ are small, we have to evaluate (9) for small $z$.

---

[2] B. v. d. Pol and H. Bremmer, Phil. Mag. 27, Eq. (10), 265 (1939).

[3] C. R. Burrows and M. C. Gray, Proc. I.R.E. 29, 14 (1941).

Hence we write

$$\frac{N(u)}{Q(u_0)}=\frac{N(u_0)}{Q(u_0)}\cdot\frac{N(u)}{N(u_0)}=[(qk_0^2)^{\frac{1}{2}}/T]$$

$$\times\{-i+[(u_0)^{\frac{1}{2}}N(u_0)-i\dot{N}(u_0)]/Q(u_0)\}$$

$$\cdot\left[1+\frac{z\dot{N}(u_0)}{N(u_0)}+\cdots\right], \quad (12)$$

$$T\equiv[(k_0^2-k^2)^{\frac{1}{2}}+\gamma(k_1^2-k^2)^{\frac{1}{2}}],$$

$$\Psi(0, z)=\phi+\psi+z\phi', \quad (13)$$

$$\phi=-2ie^{i\omega t}\int_0^\infty J_0(kr)kdk/T,$$

$$\psi=2e^{i\omega t}\int_0^\infty J_0(kr)kdk$$

$$\times[(u_0)^{\frac{1}{2}}N(u_0)-i\dot{N}(u_0)]/Q(u_0)T, \quad (14)$$

$$\phi'\simeq-2ie^{i\omega t}\int_0^\infty J_0(kr)kdk\dot{N}(u_0)/N(u_0)T. \quad (15)$$

In (15) we have not included the contribution from the second term in the braces of (12) because it is of the order of $1/\beta$ relative to the contribution from the first term. The integral in (15) is to be understood as the limit approached as $h\rightarrow 0$ when a factor $\exp[-ih(k_0^2-k^2)^{\frac{1}{2}}]$ is inserted in the integrand. Since for large $\beta$, $Q\simeq\beta N(u_0)$ we have

$$\psi=(\Lambda+\phi')/\beta,$$

$$\Lambda\equiv2e^{i\omega t}\int_0^\infty J_0(kr)kdk(u_0)^{\frac{1}{2}}/T. \quad (16)$$

It can be shown that for large $\beta$

$$\Lambda\rightarrow-Le^{i\theta}/(\beta r^2)=-\beta\phi,$$

$$\theta=(\pi/2)+\omega t-k_0r. \quad (17)$$

Hence

$$\phi'=\beta(\phi+\psi)=\beta\Psi(0, 0);$$
$$\Psi(0, z)=(1+\beta z)\Psi(0, 0), \quad (18)$$

where $\Psi(0, 0)$, the surface wave on a spherical earth, is given by

$$\Psi(0, 0)=(e^{i\theta}L/\beta^2r^2)g(p'). \quad (19)$$

Combining (18) with (8) we arrive at an expression for the field in terms of the surface wave when both $z_1$ and $z$ are small:

$$\Psi(z_1, z)=(1+\beta z_1)(1+\beta z)\Psi(0, 0). \quad (20)$$

In Fig. 1 a comparison is shown between the field obtained by van d. Pol and Bremmer[2] and values computed from Eqs. (20) and (19). Here $z_1$ or $z$ is not more than 0.63, so that Eqs. (A) and (B) of I, which are valid only for $z>1$, are inadequate.

# Letters to the Editor

## Grain Growth in Alpha-Brass

J. E. Burke
*Institute for the Study of Metals, University of Chicago, Chicago, Illinois*
June 30, 1947

An extensive series of data on isothermal grain growth in alpha-brass has recently been published by Walker.[1] Plotting from these data log$D$ (grain diameter in mm), *vs.* log time, one obtains a series of isothermal curves. An example for 42.4 percent deformation prior to recrystallization is shown in Fig. 1. One should like to know whether time and temperature for this process are related by a heat of activation $Q$. If so, times at temperature $T_1$ should be related to times at temperature $T_2$ by the equation:

$$\ln(t_1/t_2) = (Q/R)(T_2^{-1} - T_1^{-1}).$$

This equation can be solved to give values of $t_1/t_2$ to convert times at a given temperature to times at any reference temperature. By taking $Q$ as 60,000 cal. per mole it is found that the data taken at all temperatures can be plotted on a smooth curve at a single reference temperature, with only a reasonable experimental scatter. Figure 2 shows this for the curves of Fig. 1, at a reference temperature of 500°C. It thus seems proper to assign a heat of activation to the process. This curve is not a straight line.

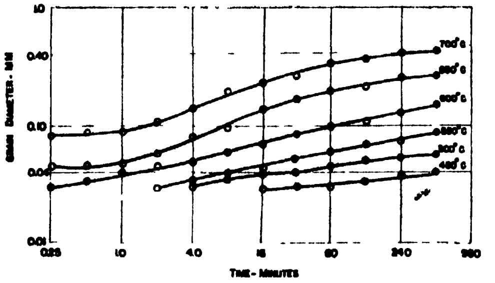


FIG. 1. Isothermal grain growth at various temperatures in alpha-brass.

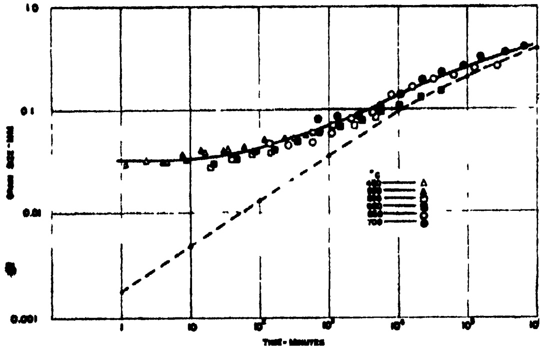


FIG. 2. Solid line: composite grain-growth curve at 500°C reference temperature; Dashed line: composite curve corrected by subtracting 0.03 mm.

Beck, Kremer, and Demer[2] report that their data for grain growth in high purity aluminum follows the relationship:

$$D = kt^n,$$

where $k$ is a constant and the exponent $n$ increases linearly with temperature. Obviously this relationship cannot hold for short times, for, at $t=0$, $D$ has a finite value, $D_0$.

The present data are expressed reasonably well by the relationship:

$$D - D_0 = k_1 t^n,$$

where $D_0$ is the grain size at the time of recrystallization, $k$ is a temperature-dependent constant, and the exponent $n$ has the value 0.425 until late in the growth process. The dotted line in Fig. 2 was obtained by subtracting the recrystallized grain size reported by Walker for this series (0.03 mm) from the solid curve through the experimental points. It can be seen that it is a straight line until the late stages of growth, but that the slope definitely falls off then. Similar results are obtained from other series of data in Walker's work.

This work was supported by the Office of Research and Inventions, U. S. N. (Contract No. N6ori-20-IV).

---

[1] Harold L. Walker, University of Illinois Eng. Exp. Sta. Bull. No. 359 (Nov. 1945).

[2] Paul A. Beck, Joseph C. Kremer, and L. Demer, Phys. Rev. 71, 555 (1947).

## Comments on "Grain Growth in Alpha-Brass"

Paul A. Beck
*University of Notre Dame, Notre Dame, Indiana*
September 12, 1947

In the above note J. E. Burke calculated a heat of activation value from grain-growth data on brass obtained by H. L. Walker.[1] Considering the scatter of the data, his value of 60 kcal./g of atom is in reasonable agreement with the heat of activation value of 73.5 kcal./g atom previously calculated for brass by Beck, Kremer, and Demer,[2] by an essentially identical method, from other published grain-growth data. Thus, Burke's result tends to confirm Beck, Kremer, and Demer's conclusions with regard to the heat of activation value from grain-growth data on brass.

Burke states that the relationship,

$$D = k \cdot t^n, \tag{1}$$

found by Beck, Kermer, and Demer for isothermal grain growth in high purity aluminum cannot hold for short times, for at $t=0$, $D$ has a finite value, $D_r$, the grain size as recrystallized. Obviously, Burke uses $t$ to denote the time for grain growth. The zero point of his time scale is at the moment when recrystallization is just complete. However, in the nomenclature used by Beck, Kremer, and Demer $t$ denotes the total time of annealing, which includes the time for recrystallization.* Clearly, Burke's objection is based on a misunderstanding of the nomenclature.

Since most of the annealing periods used by Beck, Kremer, and Demer are very long in comparison with the time for recrystallization, the error committed by using "grain-growth time" instead of "total time," as a result of the above misunderstanding, is negligible, except for the shortest annealing periods at the lowest annealing temperatures. Consequently, the whole matter would be lacking in significance were it not for another mistake in Burke's note. He uses the time data from Walker's paper as if these meant "time for grain growth" in the above sense. Actually, Walker, like Beck, Kremer, and Demer, uses the "total time" of annealing. As some of the times of recrystallization in Walker's work are relatively long (such as 16 min. at 450°C), the error resulting from this misinterpretation is significant.

The dotted line in Burke's Fig. 2 gives $\log(D-D_r)$ as a function of $\log t$, where $t$ is supposed to be the time for grain growth, but it is actually the total time of annealing of Walker's specimens, as explained above. Although the initial portion of this line was so selected across the "reasonable scatter" of the transposed experimental points that it is straight, Burke admits that deviation from the straight line definitely occurs at longer annealing periods. Nevertheless, he considers that the data are reasonably well expressed by the relationship:

$$D-D_r=k_1t^n. \tag{2}$$

A comparison of relationships (1) and (2), by using Beck, Kremer, and Demer's grain-growth data on high purity aluminum at 400°C, is given in Fig. 1. It is clear that the plot of $\log D$ *vs.* $\log t$ closely approximates a straight line. Plotting the same data as $D-D_r$ *vs.* $t_g$ ($t_g$ is used here to designate "grain-growth time" in order to avoid confusion) results in a curve which asymptotically approaches the above straight line at very large $t_g$ values. Similar curves are obtained for other annealing temperatures. Obviously, relation (2) does not fit the grain-growth data for high purity aluminum, but formula (1) does.

It can be easily shown that the initial portion of the curve in the figure approaches a slope of $n=1$. If $t_g$ approaches zero, $D/D_r$ approaches 1, that is $D/D_r=1+d$, where $d$ is a very small number, and

$$D-D_r=d\cdot D_r. \tag{3}$$

In Eq. (1) the constant $k$ can be eliminated if one puts $D=D_r$ for $t=R$, where $R$ is the time for just complete recrystallization. The following formula is then obtained:

$$D/D_r=(t/R)^n$$

or, substituting $t_g=t-R$,

$$D/D_r=(1+t_g/R)^n. \tag{4}$$

From (4), with the above relations,

$$1+d=(1+t_g/R)^n,$$

and from this, for a small value of $t_g$,

$$d=(n/R)t_g. \tag{5}$$

By substituting (5) into (3) one obtains

$$D-D_r=n(D_r/R)t_g. \tag{6}$$

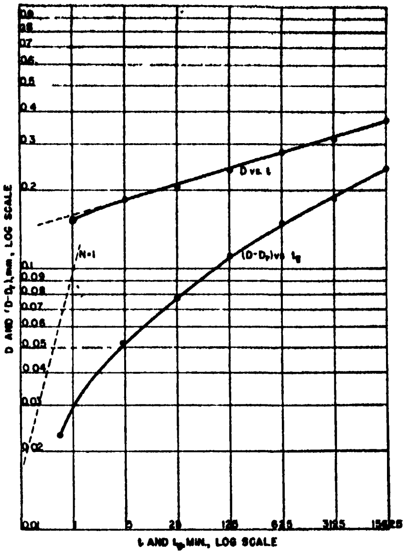


Fig. 1.

Formula (6) shows that for small values of $t_g$ the relationship between $D-D_r$ and $t_g$ is linear. This means that at very short times in the logarithmic plot the curves for all temperatures approach the slope of $n=1$. At long annealing times they approach the $D=k\cdot t^n$ straight lines, where the slope, $n$, increases with the temperature.*

* A fuller discussion of these questions is included in a detailed paper by Beck, Kremer, Demer, and Holzworth to be published in Metals Technology, September, 1947.

[1] Harold L. Walker, University of Illinois Eng. Exp. Sta. Bull. No. 359 (1945).

[2] Paul A. Beck, Joseph C. Kremer, and L. Demer, Phys. Rev. 71, 555 (1947).

## Electron Microscope and Electron-Diffraction Study of Slip in Metal Crystals

R. D. Heidenreich and W. Shockley
*Bell Telephone Laboratories, Inc., Murray Hill, New Jersey*

August 20, 1947

The process of plastic deformation in metal crystals has been the subject of many investigations, both experimental and theoretical, over a long period of years. The new experimental results to be described briefly in this letter have been presented orally before two different groups* but have not appeared in print. A fuller account of the methods and results will appear in a British publication.

The current experiments concerning the structure of slip bands were carried out using 99.99 percent aluminum cast in the form of small single crystals (20×6×2 mm). The as-cast crystals were mechanically polished through 4/0 French emery to produce a flat, 6×20 mm surface and then electropolished to remove the heavily worked layer. The samples were then annealed 3–4 hours above 600°C, electropolished again, and then etched and rinsed[1] to produce a clean, dry surface exhibiting etch pits indicating the orientation of the crystal. Electron-diffraction reflection patterns from the crystals prepared in this manner were Kikuchi line patterns: taken as a criterion for a reasonably perfect crystal.

The electron micrographs were obtained from anodic-oxide replicas of the crystals produced by anodizing in a 2.5 percent boric acid solution neutralized with ammonium hydroxide. These replicas are capable of a resolution of real structure down to at least 50A.

Plastic deformation was accomplished either by pulling the crystals in tension or by compression along the long axis of the crystal after the final surface was prepared. Examination of the surfaces with a light microscope at 50–100× indicated the orientation of the slip lines and their relative density and distribution for various amounts of plastic deformation.

The following conclusions were arrived at from combined electron microscope and electron-diffraction results on numerous aluminum crystals with various amounts of deformation.

(*A*) *Electron Microscope*—The electron micrographs clearly show that the slip bands observed in a light microscope are actually laminar regions (see Fig. 1). Observations of slip bands at various orientations to the prepared surface and their traces through etch pits indicate that the laminae are about 200A thick and of rather indefinite extent. From micrographs taken of crystals with various amounts of plastic deformation, the development of slip bands is proposed to be as shown in Fig. 2. The initial slip occurs in a single (111) plane (in the slip direction) until a displacement of about 2000A is experienced. This particular plane apparently then offers sufficient resistance to further motion that slip begins anew on another (111) plane about 200A removed from the original. The process repeats and results in a laminar slip region. In the case of two adjacent slip regions, the laminae finally merge at high deformations to produce a large field of laminae as shown in the final stage in Fig. 2.

(*B*) *Electron Diffraction*—As mentioned, electron-diffraction patterns from the newly prepared crystal are Kikuchi line patterns.

Patterns taken with the electron beam parallel to and normal to the slip traces observed on the prepared surface, at various deformations, indicated that the material between visible slip bands is relatively unaffected at small deformations. That is, the Kikuchi lines persist until the spacing of the slip bands is reduced to about 5$\mu$ or somewhat less (about 5 percent elongation). At about 5 percent elongation the Kikuchi lines disappear, leaving only a spot pattern, and at still higher deformations the spots begin to split up, indicating rotation of various parts of the crystal. Slip on a second set of planes is usually observed at elongations in excess of about 5 percent.

Patterns taken normal to the slip traces generally do not exhibit Kikuchi lines but show, instead, multiple Laue spots. As the deformation increases more spots appear, and their total angular spread increases to as much as 8–10° at 10 percent elongation.

The electron microscope and electron-diffraction results are consistent in the conclusion that practically all the distortion occurs in the slip bands and very little in the remainder of the crystal.** The laminae in a slip region experience not only a translation in the slip direction, but also slight rotation relative to one another as indicated by the electron-diffraction patterns taken normal to the traces. This rotation is probably a result of constraints imposed during deformation. It is suggested that the rotation of

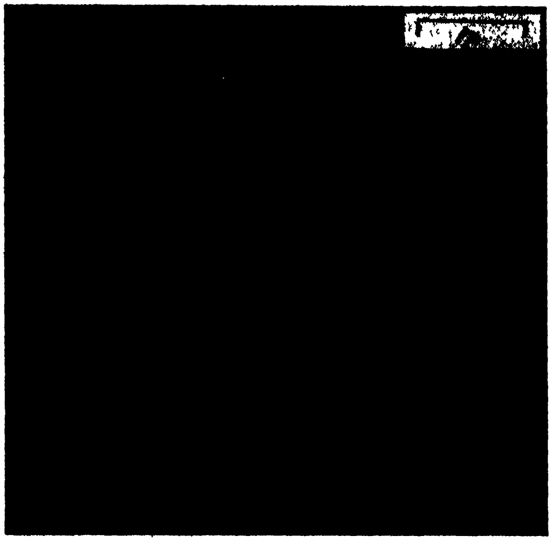


**Fig. 1. Intersecting slip bands at surface of electropolished and etched aluminum crystal showing laminar nature of slip region.**

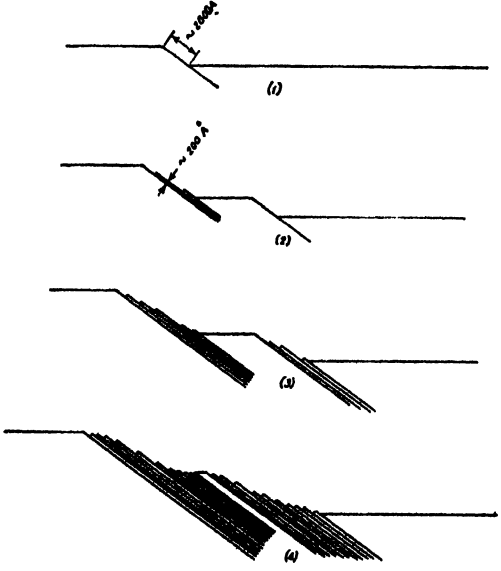


**Fig. 2. Progressive formation of laminar slip region with increasing deformation. Adjacent slip regions merge to produce large field of laminae at high deformation.**

the laminae is responsible for asterism in x-ray Laue photographs of deformed metals.

The laminar nature of the slip bands in aluminum suggests similar studies of other metals. First results show that nickel exhibits similar phenomena. In a private discussion, Mr. R. Maddin of Yale University has indicated that the results he has obtained with brass crystals are quite similar to those with aluminum.

[*] American Society for X-Ray and Electron Diffraction, Montreal, June 23–26, 1947; Bristol Conference on Theoretical Physics, Bristol, England, July 2–9, 1947.

[1] R. D. Heidenreich and L. Sturkey, J. App. Phys. 17, 127 (1946).

[**] The conclusion is in general agreement with that of Guinier from studies of the x-ray background scattering produced by a deformed aluminum crystal. A. J. Guinier, Proc. Phys. Soc. (London) 57, 310 (1945).

---

# New Books

## Helium

By W. H. Keesom. Pp. 494, Figs. 258, 25×17 cm. Elsevier Publishing Company, Inc., New York, 1942. Price $10.00.

Professor Keesom has used the war years in Holland to good advantage for the world of science by writing an excellent book about the helium atom in its gaseous, liquid, and solid states. There is, in addition, a review of work on the helium nucleus and on subjects related to the outer electrons. By compiling the experimental work on liquid helium and by inserting his keen criticism of the work, Professor Keesom has rendered in this section alone a valuable contribution to research workers in the field of low temperatures. From experience this reviewer has found that *Helium* is an active and authoritative text used by low temperature research laboratories.

The manuscript had been completed by Keesom in 1941, and for this reason the very important researches on liquid helium by P. Kapitza and by L. Landau got into the text through the supplementary notes at the end of the book. Since the publication of *Helium* in 1942, there have been published experimental and theoretical papers (again mostly by the Russians: Kapitza, Peshkov, Andronikashvilli, Lifshitz, and Landau) which would amply warrant another monograph on the subject of liquid helium by some recognized authority. Keesom has done a good job of summarizing F. London's theory about He *II* being a condensation in momentum space and the theoretical research by the Dutch workers Bijl, DeBeer, and Michels very properly receives attention. The experiments on flow of liquid helium in phase *II* are sometimes confusing, and the reader would do well to mentally sort out these as problems to be tackled by a theory after the first-order physical properties are understood. The x-ray evidence for spatial order in He *II* by Keesom and Taconis, and the subsequent theory involving holes in the lattice, is not too convincing.

Professor Keesom's book serves as an excellent source of information concerning the temperature scale, and sharply brings to attention the confusion which exists about the vapor pressure *vs.* temperature for liquid helium below 2°K. It would seem that the Leiden and Oxford laboratories have their respective viewpoints about the temperature scale below 2°K clearly stated in Keesom's book and that the eighth International Conference of Weights and Measures would render a service to the now numerous low temperature laboratories by asking its own Advisory Committee for Thermometry to use Keesom's book as a report and to recommend one temperature scale to be used by all laboratories until such time as another scale shall be warrented. The problem of an international temperature scale below −190°C will undoubtedly be taken up by the Advisory Committee for Thermometry, and in this respect Keesom's book gives prominence to the Leiden helium thermometer and the virial coefficients or gas-law correction terms used at Leiden. The Berlin helium thermometer and gas-law correction terms used by Henning have been presented by Keesom without criticism; the Berlin scale is not in precise accord with the Leiden scale.

Keesom's book is two things: (a) it is a complete source book for information about helium in its gaseous, liquid, and solid state, and (b) it is a monument to the lifetime research of Kamerlingh Onnes, of W. H. Keesom, and of their many collaborators. There are many excellent diagrams of apparatus, but space did not permit the book to reveal the intricate low temperature techniques which have made Leiden the great laboratory in this field of science.

Charles F. Squire
*The Rice Institute*

---

## Introduction to Aerodynamics of a Compressible Fluid

By Hans Wolfgang Liepmann and Allen E. Puckett. Pp. 262, 6×9½ in. John Wiley and Sons, Inc., New York, and Chapman and Hall, Limited, London, 1947. Price $4.00.

This book is a compilation in a systematic manner of most non-classified material on this extremely important phase of aerodynamics. It has been presented by two well-known authorities in this field as an introduction to the subject rather than as a standard classroom textbook, no problems and few practical examples being presented.

The book, one of the Galcit Aeronautical Series, is divided into two phases. Part I furnishes the reader a background of fundamentals, including basic thermodynamics, one-dimensional equations of motion, isentropic channel flow and applications, shock waves, and optical methods of flow observation. The chapter on optical methods is extremely descriptive, being quite brief in its mathematical discussions, although the essential introductory material is presented. The lack of illustrated examples, illustrations of flow phenomena, and comparison with experimental data, primarily caused by wartime restrictions, is a great handicap in the clarification of this section.

Part II is an introduction to the more mathematical literature on the subject and deals with two- and three-dimensional motions of a compressible fluid. This comprises an introduction to practically all present day methods of solution of compressibility problems, covering specifically the method of small perturbations, the method of development of the potential function into a power series in the thickness parameter or in the Mach number, the hodograph methods, the method of characteristics, and certain exact solutions. A very brief discussion of viscosity effects and experimental data is included, the latter being necessarily brief because of the small amount of unclassified material available in this field. Again much is lost in the lack of practical examples, comparison with experiment, and the lack of pictorial illustrations of flow phenomena. This is greatly offset by the clarity of mathematical presentation.

The book is primarily aimed for graduate consumption, requiring a good working knowledge of differential and integral calculus as well as a knowledge of advanced calculus.

A brief table of references is also included.

Sydney D. Black
*Case Institute of Technology*

---

## Antennae, an Introduction to Their Theory

By J. Aharoni. Pp. 265, 6¼×9¾ in., Figs. 149. Oxford University Press, London, 1947. Price $8.50.

The study of antennae is peculiarly difficult both theoretically and experimentally. While much progress has been made during the past decade, many interesting and important problems remain unsolved. Accordingly, all who are engaged in research in this field, and especially those who teach courses in it in graduate schools, will welcome J. Aharoni's systematic and comprehensive introduction to its mathematical theory. Because of the advanced nature of the subject, the book is, in effect, an integrated presentation of a special branch of mathematical physics. It is carefully planned, well organized, and its principal contents are delineated in sufficient detail to permit analytical sequences to be followed readily by a reader adequately prepared in the methods of classical theoretical physics and in electromagnetic theory.

The book begins with an illuminating introduction which summarizes and describes briefly the three general methods which have been used by various investigators in the theoretical analysis of antennae. Appropriately, the author uses these methods as the subjects of his three chapters: I. "Antennae and Boundary-Value Problems" (86 pages); II. "Antennae and Integral Equations" (149 pages); III. "Antennae as Wave Guides" (29 pages). The presentation of antenna theory in this manner is an important feature of the book since it emphasizes a unity of subject and of purpose in a diversity of method.

In Chapter I the introduction is followed by an excellent, concise review of electromagnetic theory. The main part of the chapter discusses with exceptional clarity typical boundary-value problems involving planes, spheres, cylinders, and cones. A study of free and forced oscillations of spheres and spheroids following the work of Stratton and Chu and of Ryder concludes the chapter.

Chapter II opens with a derivation of ordinary electric-circuit theory from general electromagnetic principles beginning with the power equation expressed in terms of the scalar and vector potentials and the densities of charge and current. This is an important and well-written section. It serves to correlate in a fundamental manner the circuit aspects of antenna theory with conventional electric-circuit analysis. This correlation is effectively pursued and developed in subsequent analyses of the magnetic dipole (or small loop) and the electric dipole. From these, the argument continues logically to thin-wire antennae for which the integral equation is derived in quite general form. The study of the center-driven antenna and of the receiving antenna follows closely the original work of Hallén and its extension by King and Blake and King and Harrison. The chapter continues with the derivation of the electromagnetic field following the conventional procedure of assuming a sinusoidal current distribution; it concludes with a section on the effect of the earth on the field, and a section on direction finding.

Chapter III is devoted to a concise but clear presentation of the theory of the bi-conical antenna of small angle and to the relationship between antenna theory and transmission-line theory as developed by Schelkunoff.

Taken as a whole, *Antennae, an Introduction to their Theory* is an authoritative and valuable book which skillfully delineates and summarizes major phases of antenna theory as published up through the year 1945. It should prove to be an excellent text and reference book for advanced students and research workers in the field of antennae and applied electromagnetism.

Ronold W. P. King
*Harvard University*

---

## Technological and Physical Investigations on Natural and Synthetic Rubbers

By A. J. Wildschut. Pp. 173+xi, Figs. 72, Tables 17, 14.5×20.5 cm. Elsevier Publishing Company, Inc., New York, 1946. Price $3.00.

This monograph is one of a series describing research in Holland in the five years of German occupation (1939 to 1944), during which period regular journal publication was interrupted. Although not intended as exhaustive, the investigations described provide a surprisingly comprehensive coverage of the field. The material is presented in the orderly manner associated with a compilatory book without neglecting those descriptions of experimental details expected in an original contribution. The monograph will be carefully studied by all scientific workers concerned with the properties of elastomers. They will find experimental results and concepts which are new, as well as those which are confirmatory of, or occasionally at variance with, material previously published elsewhere. As an example of the former, measurements of the tensile strength of unelongated and elongated rubbers at the temperature of liquid hydrogen are reported for the first time, to the reviewer's knowledge. The latter is illustrated by the results and conclusions of J. M. Goppel on the degree of crystallization as a function of elongation, measured by means of x-rays, which differ from those obtained in America by J. E. Field. It is regrettable that outside contemporary literature, particularly American and British, was for the most part unavailable during the period of preparation, precluding its integration with the work presented. However, the desirability of prompt publication, in the reviewer's judgment, outweighed any possible gain to be derived from the inspection of all published material.

Part I of the monograph is concerned primarily with technological investigations. In Chapter 1, a survey of the various rubbers and rubber-like materials investigated, their classification and formulation, is followed by data, with discussion, of a comprehensive list of mechanical properties. Then aging, electrical and standard physical properties, and resistance to chemicals are presented. Chapter 2 includes measurements on the plasticity of unvulcanized rubbers, a discussion of aging tests and the mechanism of aging, and mathematical interpretations of the stress-strain curve. Studies of the vulcanization process are given in Chapter 3, with emphasis on the use of synthetic resins as vulcanizing agents.

Part II is devoted to more fundamental physical studies. It is prefaced, in Chapter 4, by a discussion of physical methods in structural research. Plastic and elastic properties, with which Chapter 5 is concerned, are suggested as the most important. The thermodynamics of vulcanized rubbers, the plastic flow, the elasticity and permanent set, and the internal friction are each analyzed in considerable detail. Experimental techniques are described. The effects of very low (−253°C) and high temperatures and of swelling on the mechanical properties, particularly tensile tensile strength, are shown in Chapter 6. The x-ray investigations previously mentioned are set forth in Chapter 7. Several investigations on the electrical properties of rubber are treated briefly in Chapter 8. Finally, in Chapter 9, general conclusions of the entire work, emphasizing the uses and limitations of the various rubbers and the advances in knowledge of fundamental properties bearing on the theory of structure, are summarized.

I. B. Prettyman
*The Firestone Tire and Rubber Company*

# New Booklets

The Gaertner Scientific Corporation, 1201 Wrightwood Avenue, Chicago 14, Illinois, announces the publication of *Bulletin 151–74, Large Two-Lens Quartz Spectrograph.* 8 pages, free on request.

Leeds and Northrup Company, 4902 Stenton Avenue, Philadelphia 44, Pennsylvania, publishes a quarterly newspaper-style bulletin entitled *Modern Precision.* 8 pages, many photographs.

Burrell Technical Supply Company, 1936 Fifth Avenue, Pittsburgh 19, Pennsylvania, has published a 4-page bulletin describing its "unit-package" tube furnaces for temperatures up to 2650°F.

Dow Corning Corporation, Midland, Michigan, announces an issue of silicone notes entitled *DC 710 Silicone Oils.* 8 pages, free on request.

Sound Products Section of Radio Corporation of America has published a comprehensive new manual on the engineering and installation of sound systems. It is designed for use by architects, building engineers, and the construction industry generally, and covers all types of buildings. 288 pages, bound in simulated black leather, fully illustrated and indexed. Available at $5 from RCA sound systems distributors or direct from the RCA Sound Products Section, Camden, New Jersey.

# Here and There

### New Appointments

David A. Katcher has been appointed editor of a new semipopular magazine on physics to be published early next year by the American Institute of Physics. The new magazine is planned to fill some of the gap which now exists between specialized technical journals and popular science magazines. Katcher comes to the Institute from the Naval Ordnance Laboratory in Washington, D. C.

Isaac L. Auerbach, Albert Auerbach, Gerald Smoliar, and Joseph D. Chapline, Jr. were recently added to the staff of the Electronic Control Company, Philadelphia, Pennsylvania.

Cyril J. Staud has been named Director of Kodak Research Laboratory, and John A. Leermakers is the new Assistant Director and head of the Photographic Theory Department.

Walter Ramberg has been appointed Chief of the Mechanics Division of the National Bureau of Standards, succeeding Hugh L. Dryden, who is leaving the Bureau to become Director of Research for the National Advisory Committee for Aeronautics. The appointment of Grote Reber, radio physicist and engineer, to the staff of the Bureau was also announced recently.

R. C. Mason has been named Manager of the Electro-Physics Department of the Westinghouse Research Laboratories, succeeding Gaylord W. Penney, who was recently appointed Westinghouse Professor of Electrical Engineering at the Carnegie Institute of Technology.

### Princeton Professor Drowned

Assistant Professor Shuichi Kusaka, a member of the Department of Physics at Princeton University and an authority on cosmic-ray theoretical research, was drowned on August 31 while swimming at Beach Haven, New Jersey.

### RCA Fellowship Award

An RCA Fellowship was established in January 1947 for the purpose of encouraging RCA engineers to obtain advanced degrees and to help in the recruiting of young engineers for research and development. First RCA Victor employee to be awarded the fellowship was Harry J. Woll, engineer at the company's Camden plant. He will study for the Ph.D. degree at the University of Pennsylvania, and during this period will be on leave of absence from the company.

### Acta Crystallographica

A new journal of crystallography, Acta Crystallographica, sponsored by the recently formed International Union of Crystallography, will start publication in January 1948. (See Journal of Applied Physics, April 1947.) It will be published by the Cambridge University Press under the editorship of Professor P. P. Ewald, a former editor of the Zeitschrift für Kristallographie, and a panel of regional co-editors. The annual subscription will be $10 for six issues and, by courtesy of the American Institute of Physics, arrangements have been made for subscribers in the U.S.A., its possessions, and Canada to subscribe through that body. Subscribers in other countries should subscribe directly through the Cambridge University Press, Bentley House, Euston Road, London, N.W. 1, England. Articles for publication will be accepted in English, French, German, or Russian, and authors in the U.S.A. are invited to submit manuscripts to Professor I. Fankuchen, Brooklyn Polytechnic Institute, 85 Livington Street, Brooklyn 2, New York.

### Westinghouse Scientist Honored

Arpad L. Nadai, Consulting Engineer of the Westinghouse Research Laboratories, has been awarded the Worcester Reed Warner Medal by the American Society of Mechanical Engineers, honoring noteworthy contributions to the permanent literature of engineering. Earlier this year Dr. Nadai delivered the James Clayton Lecture before the British Institution of Mechanical Engineers in London and Manchester.

### British American Research, Limited

National Research Corporation, Cambridge, Massachusetts, announces the formation of British American Research, Limited, which has recently been incorporated in Scotland as a joint venture with Daniel Varney, Limited. The latter company was engaged during the war in the manufacture of scientific research instruments and laboratory apparatus. The new company will undertake the manufacture and sale of high vacuum equipment.

Journal
of
Applied Physics

Volume 18, Number 12 December, 1947

# A Preliminary Study of a Physical Basis of Bird Navigation

HENRY L. YEAGLEY
*Department of Physics, The Pennsylvania State College, State College, Pennsylvania*

(Received June 26, 1947)

## Foreword

F. C. WHITMORE
*Dean of the School of Chemistry and Physics, The Pennsylvania State College*

The theory presented was formulated in 1942 at the Pennsylvania State College by Dr. H. L. Yeagley, Associate Professor of Physics. Its only public presentation is in a preliminary non-technical article in the October 1943 issue of The Penn State Engineer, a local student publication. The training of birds for the first experiments was begun in May 1943. The first actual field test was made November 7, 1943 with homing pigeons trained to navigate to Paoli, Pennsylvania. This experiment was witnessed by officers of the U. S. Army Signal Corps, including Major Otto Meyer, in charge of the Army Pigeon Service Agency of the U. S. Army Signal Corps. Because of the positive nature of the results and their importance to communication in the war, all subsequent work was reported only to the Army Service Forces:

Report on Bird Navigation, "Theory and confirming evidence" Decemder 12, 1944.
Report on Bird Navigation, "Hamlet experiment," April 24, 1945.
Report on Bird Navigation, "A continuation of experiments started at the Pennsylvania State College in December, 1942," January 4, 1946.
Report on Bird Navigation, "East-West experiment," June 30, 1946.

The bird navigation theory as presented implies an organ or organs in the bird's physiology which are sensitive to the effect of its motion through the vertical component of the earth's magnetic field and to the effort exerted to overcome the coriolis force, due to the earth's rotation. Both these influences involve a set of lines which together form a navigational gridwork. By correlating its instantaneous land speed with the two above effects, a bird can fly to its home which is a unique point in this gridwork, or to related companion points existing in the gridwork at positions other than its home. Experimentation with homing pigeons between the home and spurious home points (conjugate or companion points) has yielded data which supports the theory.

PROBABLY no natural phenomenon has intrigued man and held his interest more than the ability of the homing pigeon to fly long distances from a strange locality and across unfamiliar territory to its home loft. Many leading ornithologists believe that this same ability guides wild birds in their long migratory flights over land and water areas. Invariably, references made to this remarkable accomplishment in the bird world are in terms of "instinct," which is a way of saying "we don't know."

The over-all problem of bird navigation involves at least the following five factors:

1. A desire on the part of the homing pigeon

to go to the home loft or, in the case of wild birds, to go to their summer feeding and breeding grounds, or their warmer wintering areas during migratory periods.

2. The physical strength and endurance to fly the required distance.

3. A knowledge of familiar landmarks, within a minimum radius, for piloting to the exact spot. (This radius distance for homing pigeons is known to be about twenty-five miles.)

4. The ability to navigate. This implies the possession and use of an organ or organs as navigating instruments.

5. Considerable navigating experience in flying in and about the "home" region. With homing pigeons this experience is artificially built up by sending the birds to gradually increasing distances.

It is well known that many birds of the migratory species return to their breeding and nesting grounds after being trapped and sent hundreds of miles into strange territory. Many such flights have been arranged so that the birds must orient themselves and fly hundreds of miles over water routes out of sight of land. One species, the Golden Plover,[1] performs the remarkable feat of navigating 3000 miles from Alaska to the Hawaiian Islands with no landmarks over the broad expanses of the Pacific Ocean. Hundreds of individuals of the albatross family roam over thousands of square miles of the Southern Pacific area. Members of each species finally return to their own tiny island to breed and raise young. Hood Island[2] of the Galapagos Archipelago group, is headquarters for one species of these birds. Since there are no "sign posts" of any kind over the ocean wastes, the flights must involve true navigation until the home island comes into view. Other types of birds, including noddy and sooty terns, swallows, herring gulls, numerous varieties of ducks, hawks, storks, starlings, chimney swifts, have the homing or navigating ability developed to a high degree.

Thus, while man has been slowly and laboriously learning to navigate during the past six centuries, birds have been using a system of their own for perhaps millions of years.

Hundreds of papers have been written suggesting possible fragmentary explanations of the mystery of bird orientation. The more scientifically possible of these explanations involve sensitivity of the bird to one or more of the following:

1. The effect of motion through the earth's magnetic field.[3]

2. Gravitational differences due to variations in the earth's radius.

3. The Coriolis effect due to the earth's rotation.[4]

In order to set up a working system of bird navigation or, indeed, of any navigation, a combination of two or more overlapping fields of influence is necessary. Each of these must possess definite gradients. Possible grid patterns might represent repetitive, displaced fields of a single kind, or combinations of two or more of the following:

1. Light beams—limited, by horizon extinction, to comparatively small areas, and useful only in piloting.

2. Sound beams—too limited for long distances, conceivably available only for piloting.

3. Electric fields—extremely limited by small intensity factor, variable magnitudes, lack of permanence.

4. Electromagnetic radiations—network of radio or radar beams such as are used in the radio-range network system of navigation or, more properly, piloting.

5. Magnetic fields—the earth's natural magnetic field, a somewhat irregular function of distance from the magnetic poles.

6. Force fields, such as:

   a. Coriolis force,
   b. Gravitational force,
   c. Centrifugal force.

---

[1] Frederick C. Lincoln, *The Migration of American Birds* (Doubleday, Doran and Company, Inc., Garden City, New York, 1936).

[2] J. L. Peters, *Check-list of Birds of the World* (Harvard University Press, Cambridge, Massachusetts, 1931), Vol. 1 or 2; *ibid.*, Vol. 2, 1937; *ibid.*, Vol. 3, 1940; *ibid.*, Vol. 4.

[3] C. Viguier, "Le seus de l'orientation," Rev. Phil. 14, 1 (1882); A. Thauzies, "Apropos d'une theorie sur l'orientation du pigeon voyageur," Rev. Sci. (series 5) 3, 270 (1905); "L'orientation du pigeon voyageur," Ann. Med. Vet. 54, 141 (1905).

[4] Dr. W. H. Thorpe and D. H. Wilkinson, "Ising's theory of bird orientation," Nature, December 21 (1946). Also, G. Ising, "The physical possibility of a biological sense of orientation based on the rotation of the earth," Ark. Math. Astrophys. 32A, No. 4 (1946).

Appended is a bibliography of the fragmentary suggestions that have previously been made in explanation of bird navigation. Special attention is called to Dr. D. R. Griffin's paper,* "The sensory basis of bird navigation," and to W. M. Levi's book, *The Pigeon*,† for an excellent general coverage of the subject.

In setting up the present theory of bird navigation the following well-known facts pertaining to the flying and use of homing pigeons were considered:

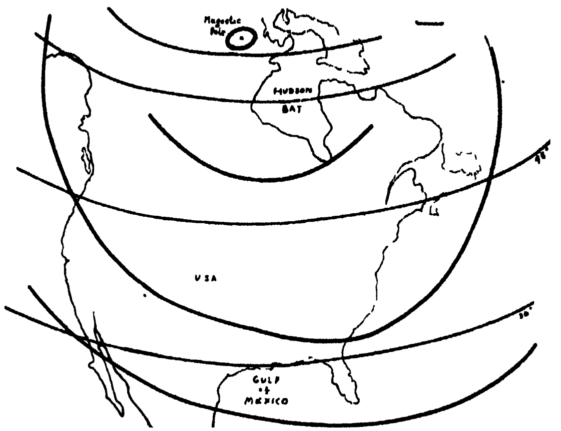
1. When released at unfamiliar and distant points, they usually fly in large circles for a matter of minutes before moving off in a straight line. When training flights have been experienced once or twice daily for weeks, and pigeons are released individually from new locations around the compass, they often sense the correct direction, after having flown only one or a fraction of a complete circle.

2. Homing pigeons are unable to navigate in a thick haze or fog or in complete darkness. Blinded or hooded birds fly first in a large circle and then in a decreasing spiral, gradually decending until about six feet above ground level, after which a "crash" landing is made.

3. Pigeons are unable to navigate successfully in winds much over thirty-five miles per hour.

4. After reaching four or five weeks of age, homing pigeons can be trained "around the compass" for six or eight weeks at increasing distances of one, two, four, eight, sixteen, thirty-two miles (repeating each distance one or two times before increasing), after which they are easily capable of navigating flights of distances up to seventy miles and more. With a few more weeks of training, flights of one to two hundred miles, or even more, are readily accomplished. A year later, with some additional short training flights added to their experience, they can readily navigate home from four hundred miles, without previous experience of any kind in the new territory.

5. Written reports by the Army and Navy Communications[5] as well as numerous oral reports by representatives of flying-clubs and individuals, indicate that pigeons are confused and are unable to orient themselves when released from points near powerful radio and radar broadcasting stations. There are some indications that the audiofrequencies rather than the carrier-frequencies are the disturbing factor.



MAP I. North American section of earth-globe showing lines of equal magnetic vertical-Coriolis gridwork. South magnetic pole displacement from north terrestrial pole is roughly about 20° or 1400 miles.

The following theory is presented as the first complete working hypothesis to explain homing-pigeon navigation.

Three major factors are involved:

1. Sensitivity of birds to the effect of flying through a magnetic field.

2. Sensitivity to the forces, produced by the earth's rotation, acting on masses moving over its surface in a straight line (part of the well-known Coriolis effect which is a function of latitude).

3. Visual sensitivity to velocity over the earth's surface (land speed).

By correlating the results of the first and last of these sensitivities the bird can detect his "magnetic latitude." This differs from his true latitude because the magnetic poles are grossly and asymmetrically displaced from the earth's poles (Map 1). By similar correlation of the second and last sensitivities it is possible to detect the true latitude.

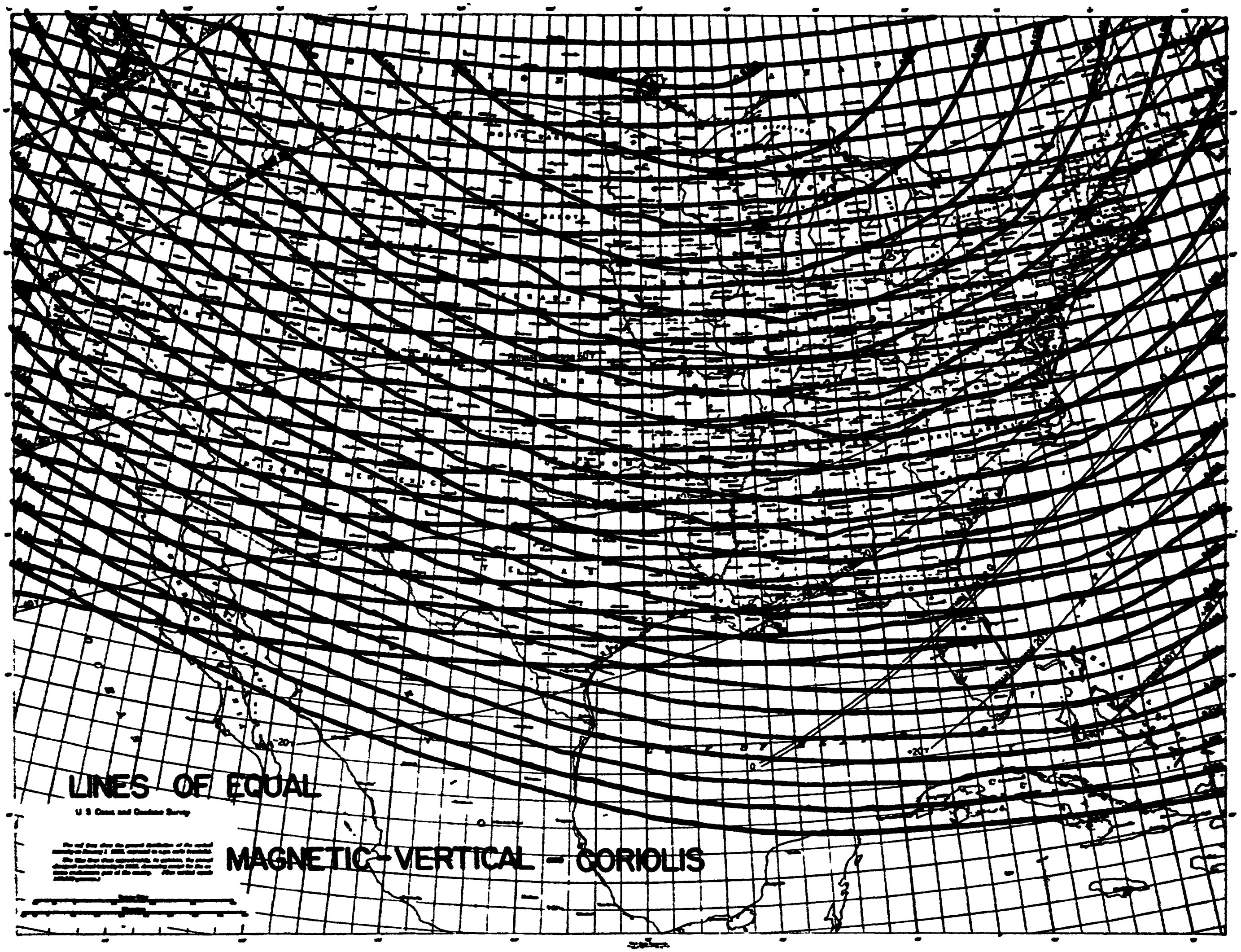
The effective result of these two correlations is that the bird can recognize his home locality at the intersection of a characteristic line in the

---

* D. R. Griffin, "The Sensory Basis of Bird Navigation," Quarterly Review of Biology, 1 (1944).

† W. M. Levi, *The Pigeon* (R. L. Bryan Company, Columbia, South Carolina, 1944).

[5] *Experiments With Homing Pigeons' Sensitivity to Radio Frequency Waves*, conducted by 285th Signal Pigeon Company, March 1945.



Map II. Lines of equal magnetic vertical-Coriolis gridwork for the United States and bordering areas.

earth's magnetic vertical field with a characteristic line of latitude.

Thus it appears that the secret of bird navigation depends on the displacement of the magnetic poles from the rotational poles of the earth with the subsequent intersection of the lines of equal magnetic-vertical intensity with latitude lines (Maps I and II).

The fact is that it was a study of Map II in relation to the failure of homing pigeons near Indianapolis, Indiana,[6] which gave the first clue to the present theory. As will be noted, in that area the two sets of curves are essentially parallel. Thus in that section, the two sets of correlations guide the birds to a line many miles long instead of to an approximate point. It should be added that the study of maps involving lines of equal magnetic declination, equal magnetic dip, equal magnetic horizontal intensity, equal total magnetic intensity, or any combination of these failed to show such parallelism in southern Indiana.

In operation, the navigation may be as follows:

1. A bird is sensitive by virtue of some organ or organs to the magnitude of the effect of its motion in flight through the vertical component of the earth's magnetic field.[7] This effect which must be correlated with its land speed, is a direct function of the bird's horizontal velocity, and the distance from the magnetic poles. It is a maximum at the poles and decreases somewhat irregularly from each pole toward a line approximately half-way between. By this means, a bird, if flying in a location of magnetic vertical field intensity different from its home, can consciously fly in a direction which will bring its land-speed magnetic vertical-field effect[8] back to that to which it is accustomed during its normal flight around home territory. Since the lines of equal vertical magnetic-field intensity are irregular closed curves centering around the magnetic poles, it is evident that they represent one set of lines involving a velocity-magnetic vertical-field-gradient pattern which is part of our proposed navigational grid system.

2. A bird is sensitive by virtue of some organ or organs to the magnitude of the Coriolis effect. This latter results from a natural relationship between the earth's rotational velocity and motion of a body over the earth. The forces involved are a direct function of the earth's rotational speed, the speed of a body moving over the earth's surface, and the latitude of the moving body. As in the case of the electromagnetic effect, the bird must correlate its land speed with the magnitude of this force in order to use it as a guide. The lines of equal Coriolis force are true circles coincident with parallels of latitude. Hence, if a bird, usually subject to the Coriolis effect of normal flying around home territory, is displaced north or south, he need only fly in a direction which will bring the "feel" of the Coriolis land-speed relation back to normal and he will be back on his own latitude circle. These circles represent a second set of lines which in bird orientation involves a force land-speed-gradient pattern which is a part of our postulated navigational grid system.

Map I illustrates the existence on the earth of a gridwork of two distinct sets of curves, representing the land speed *vs.* electromagnetic effect and the land speed *vs.* Coriolis effect. For any horizontally moving body over the earth's surface these intersecting curves give rise to a system of unique pairs of conjugate points.

This conception offered an opportunity for the experimental testing of the theory. If a bird locates its home by navigating to the intersection of the proper magnetic vertical and the proper latitude, he should as readily navigate to the conjugate point as to his home, if released nearer the former. As can be seen on Map II the conjugate point of State College, Pennsylvania is in central Nebraska. A location near Kearney, Nebraska, has been used for checking the theory, as will be described below.

## EFFECT OF MOTION THROUGH THE MAGNETIC FIELD

It is a well-known fact that whenever a continuous ring or circle of some material medium is rotated in a magnetic field, so that the lines of flux are cut, an e.m.f. is induced. This is said to result from a change of flux linkage and its magnitude is given by:

---

[6] F. J. Sauerteig, "Indianapolis," Am. Racing Pigeon News, Nov. 1942, p. 1.

[7] See reports to the U. S. Army Signal Corps listed in the Foreword.

[8] This effect may be the same phenomenon as the so-called "electromagnetic effect," it will subsequently be referred to as such for the sake of brevity.

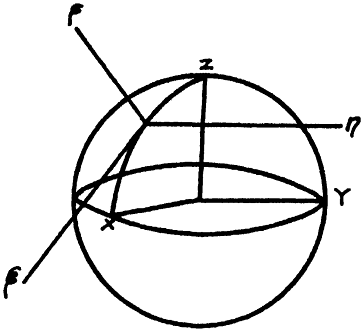


FIG. 1.

e.m.f. $= d\psi/10^8 dt$ (e.m.f. = induced voltage);

$d\psi/dt$ = rate cutting of flux in number of lines cut per second;

$10^8$ = factor needed to give result in volts or joules per coulomb.

Although it is quite customary in physics texts to give problems in which it is required to find the voltage induced when a straight wire of given length cuts a magnetic field of given strength and orientation at a given velocity, no known experiment has ever shown conclusively that an e.m.f. is induced under these latter conditions. Nevertheless, in dealing with bird navigational flight in this theory the assumption is made that the bird in some way senses the effect of flying normally through the vertical component of the earth's magnetic field, and the laws obeyed are the same or analogous to the electromagnetic effect. Whether or not it is the latter, as is ordinarily meant in the physical sense, or a phenomenon for which man has not yet conceived a picture, does not alter its general application to the theory involved. For if a bird navigates according to a pattern which fits exactly into the vertical, magnetic-field Coriolis gridwork, illustrated in Map II, some mechanism relating it to bird navigation exists, even though man does not yet comprehend its full import. A flying bird or any object moving at right angles to a magnetic field or its component would constitute a system comparable to a moving straight wire mentioned above rather than a rotating ring.

Should the sensitivity of the bird involve the ordinary laws of electromagnetism, the magnitude of the induced voltage for a bird flying forty miles per hour at State College, Pennsylvania, where the vertical component of magnetic field intensity is .556 oersted,[9] would be:

$$\text{e.m.f.} = \frac{d\psi}{10^8 dt} \quad \frac{H \cdot L\, ds}{10^8\ dt} \quad \frac{.556 \times 1 \times 1798}{10^8}$$

$10^{-5}$ or 10 microvolts per cm of length.

$d\psi$ = lines of flux cut, $H$ = field intensity, $L$ = length of conductor, 40 mi./hr. = 59 ft./sec. or 1798 cm/sec. This voltage might make itself felt either across the nerves in the wings and other gross parts of the bird or across the eye, the ear, or any organ or parts of organs.

Not only sensitivity to the absolute value but also reaction to its variation is necessary to guide the bird in the compensatory direction. For this reason the increments of magnetic vertical-field intensity should be examined.

Average values of the vertical component of magnetic-field intensity are 0.60 oersted at a distance of 15° from the south magnetic pole in the Hudson Bay region, and 0.68 oersted at a distance of approximately 5 to 8 degrees from the South Magnetic Pole in the general region south of New Zealand. These values diminish to zero at points along great circles, approximately midway between. In equatorial regions the variation of the vertical intensity in a general north-south direction is about .01 oersted per degree. In the United States the corresponding variation is roughly .007 oersted per degree. Since all experimental work involved in this report has been done in this country the latter value will be considered for illustrative purposes.

$$\frac{d(\text{e.m.f.})}{d\varphi} = \frac{L}{10^8}\,\frac{ds}{dt}\,\frac{dH}{d\varphi} = \frac{1.798 \times 10 \times 7 \times 10^{-3}}{10^8}$$

$$= \frac{1 \times 1798 \times .007}{10^8} = \frac{1 \times 1.798 \times 7}{10^8} = 12.6 \times 10^{-8}$$

= .13 microvolt change for approximately each degree or seventy miles of flight (a change of 1.3 percent per degree).

$d\varphi$ = increment of "magnetic latitude"

---

[9] World values of $H$ may be obtained from map No. 1702, "The vertical intensity of the earth's magnetic force," published by the Hydrographic Office, Washington, D. C. Values for the United States of America from Booklet and Serial number 602, as previously indicated.

## EFFECT OF MOTION THROUGH THE CORIOLIS FIELD

The following equations represent the Coriolis acceleration or force per unit mass and were obtained from A. G. Webster's *Dynamics* (G. E. Stechert and Company, New York), p. 320–322.

$$d^2\xi/dt^2=2\Omega\sin\varphi(d\eta/dt),\quad \text{(see Fig. 1)} \tag{1}$$

$$d^2\eta/dt^2=-2\Omega\{(d\xi/dt)\sin\varphi+(d\zeta/dt)\cos\varphi\}, \tag{2}$$

$$d^2\zeta/dt^2=-g+2\Omega\cos\varphi(d\eta/dt). \tag{3}$$

$\varphi$ = latitude of place, $\Omega$ = angular rotational speed of the earth, $-g$ = acceleration of gravity, $\pm\eta$ = distance moved east or west on the earth surface, $\pm\zeta$ = distance moved north or south on the earth surface, $\pm\xi$ = distance moved toward zenith or nadir, $t$ = time.

According to these equations, it is apparent that a bird might conceivably make use of Coriolis forces due to its motion in any direction (i.e., horizontal, vertical, or their resultant directions) as part of its system of navigation. Only horizontal motions are considered as part of this theory but if vertical motions were also included, no basic changes would be involved in the discussion, since in any case the magnitude of the resulting Coriolis forces will still be a function only of $\Omega$, $\sin\varphi$ and $ds/dt$. Since the bird's continuous and normal motion in homing is horizontal, only the latter will be considered. Hence the following modified equation is obtained by combining (1) and (2):

$$\begin{aligned} a&=[(d^2\xi/dt^2)^2+(d^2\eta/dt^2)^2]^{\frac{1}{2}}\\ &=2\Omega\sin\varphi[(d\eta/dt)^2+(d\xi/dt)^2]^{\frac{1}{2}}. \qquad (4)\\ &=2\Omega\sin\varphi(ds/dt) \end{aligned}$$

In (2) the term $(d\zeta/dt)\cos\varphi$ is zero since motion in the $\zeta$ (vertical) direction is considered zero. Equation (3) denotes that there will be a vertical force of Coriolis resulting whenever there is a component of horizontal velocity eastward or westward. When the motion is in the $+\eta$ (eastward) direction it is downward, and *vice versa*. The value of $d^2\zeta/dt^2$ (insofar as the bird is concerned) varies above or below $-g$ when it flies in any horizontal direction having a component of velocity either east or west. This vertical force could also be part of the bird's orienting mechanism and might operate to help maintain migrating birds on a true north or south course. In the bird's home territory where it recognizes local landmarks and awareness of direction is a factor, as in piloting operations, this vertical force may again operate in guiding him. As previously stated, even though this vertical force, or the term $(d\zeta/dt)\cos\varphi$ (in Eq. (2)), actually enters the picture, it in no way changes the over-all concept of the Coriolis land-speed relation to bird navigation as set forth in the theory. As homing pigeons are known to fly at rates of twenty to sixty miles per hour, a value of forty miles per hour, or fifty-nine feet per second, will again be used for making an illustrative calculation. A median value of 45° latitude will also be chosen arbitrarily to go with this arbitrary speed value. Substituting in Eq. (4):

$$a=2\times\frac{2\pi}{86{,}400}\times.707\times59$$

$$=\frac{4\times3.14\times7.07\times5.9}{8.64\times10^4}=\frac{524}{8.64\times10^4}=61\times10^{-4}$$

$$=.0061 \text{ feet per second per second.}$$

The direction of this Coriolis acceleration on the northern hemisphere is to the right and on the southern hemisphere to the left of the line of flight. This latter is commonly known to produce the counter-clockwise and clockwise whirlpools of moving fluids on the northern and southern hemispheres of the earth.

As a result of this accleration, the force per unit mass would be $F=\text{ma}=1\times.0061$ or .0061 poundal per pound at 45° latitude, and a speed of forty miles per hour. Or in terms of rate change of Coriolis effect with respect to time:

$$\frac{da}{dt}=2\Omega\frac{ds}{dt}\cos\varphi\frac{d\phi}{dt}\cdot$$

But

$$\frac{d\phi}{dt}=\frac{1}{r}\cdot\frac{ds}{dt},$$

$$\therefore\frac{da}{dt}=\frac{2\Omega\cos\varphi(ds/dt)^2}{r}=\frac{2\times2\pi\times.707\times59\times59}{86400\times3955\times5280}$$

$$=\frac{4\times3.14\times7.07\times5.9\times5.9\times10^1}{8.64\times3.96\times5.28\times10^{10}}$$

$$=1.72\times10^{-8} \text{ ft./sec./sec. each second,}$$

or about

$$10.8\times10^{-5}=.000108\frac{\text{ft./sec./sec.}}{\text{degree}}.$$

This 1.7 percent change of force per unit mass therefore occurs in flying one degree normal to the parallel circles of equal Coriolis effect at 45° latitude. The absolute value of Coriolis acceleration varies from maximum $1.00/.707\times.0060=.0085$ ft./sec./sec. at forty mi./hr. at the poles to zero at the equator.

Judging by these tiny increments of Coriolis force and magnetic effect which the bird would have to correlate with land speed for navigational guidance,[4] the tendency is to reject its validity.

Also it is certain that there is a physical basis for this ability to navigate which pigeons constantly demonstrate. It seems fair to assert that if the directives involved were of a gross magnitude, with modern scientific tools and mathematical means of investigation they would have been uncovered before now.

Examination of Maps I and II reveals that in the equal magnetic-vertical intensity Coriolis-force gridwork each pair of intersecting lines cross each other at least twice. In addition, about midway between the pairs of conjugate intersections, approximate parallelism often exists over a considerable portion of their lengths. These conditions also differ from a perfect pattern because large segregated surface deposits of magnetic materials on the earth prevent the equal magnetic-vertical lines from more than simulating the form of a true circle. On the United States Map II it can be seen that this condition of parallelism obtains over a fairly wide band of land area, including parts of Wisconsin, Illinois, Indiana, Kentucky, Tennessee, etc.

The above-mentioned relationships suggest that birds trained to navigate to one point of intersection of two lines toward the east, should, if released nearer to their conjugate point, fly preferentially toward the latter. Likewise, the region of parallelism or tangency should be one of confusion for navigating birds. If trained within this region their home would not contain a unique point with reference to the theory, or, if trained outside but released within this region, they would be in the dilemma of being apparently able to fly toward their home in either of two directions, approximately 180° apart. This suggests the experiments of training homing pigeons to navigate to one location and releasing them near the conjugate point of the location, or at points in between the two conjugates to see if they react to the home or conjugate positions.

TABLE I.

| Nov. | Day | Two copper plates | Two magnets | Lost one magnet | Lost two magnets |
|---|---|---|---|---|---|
| 7 | 1st | 5 | – | 1 | – |
| 8 | 2nd | 3 | – | – | |
| 9 | 3rd | – | – | – | - |
| 10 | 4th | – | 1 | 2 | 2 |

Another suggested experiment arises from the idea that if birds are guided by the effect of flight through the vertical component of the earth's magnetic field, they should become confused if a sufficiently strong superimposed magnetic field moves upward and downward across their bodies simultaneously. These and the previously mentioned considerations led to the following set of experiments which will be described in chronological order.

## EXPERIMENT I

### The Magnetic Wing Experiment, November 7, 1943

The first experimental test of the theory was made by superimposing on homing pigeons while in navigational flight, a magnetic field moved up and down by their wing motion. This was accomplished by attaching hyflux-chrome magnets on the under side of the wings, between the first and second joints.

The normal magnetic effect would be the one due to the somewhat pulsating forward motion at right angles through the vertical component of the earth's magnetic field. If a number of homing pigeons with the superimposed moving experimental field were unable to home normally and a like number with identical training but with attached copper plates of the same approximate mass and size should home from the same release point in the normal fashion, it would be an indication that the earth's magnetic field plays a part in bird navigational guidance.

The birds used in this experiment were raised and trained for it by Mr. Lowell Gable of Paoli, Pennsylvania. Because he is an educated man with a scientific mind and one of the foremost in the field of homing pigeon racing, it was felt the experiment would be more significant than under other circumstances. Mr. Gable carefully trained twenty of his best young birds to pilot and navigate to his loft from distances up to one-hundred miles. The training flights were always from stations to the west, northwest, and north of Paoli so that the birds would not become familiar with land areas to the south and southwest over which the final experimental flights would be run. They were singly released many times at fifty to one-hundred airline miles or more to insure individual navigating skill and eliminate those that simply followed their loft-mates home. Between 10.05 A.M. and 12:04 P.M. on the morning of November 7, 1943, the twenty trained birds were released at a point eight miles southeast of Bellaire, Maryland. The airline distance to the loft at Paoli was about sixty-five miles. To the wings of each bird in turn was attached either a pair of magnetic or copper plates before it was set free. None was released until the previous one had been out of sight for five minutes. Its general direction of travel at time of disappearance was carefully noted. All even-numbered birds carried hyflux-chrome magnets, whereas the odd-numbered ones had copper plates attached to their wings. Both magnets and copper plates used were 1″×.218″×.025″ and weighed approximately .8 gram. Attachments were made on the under side of the manus portions of the wings. Since any attachment involving the wing feathers or feather follicles would have caused rapid and serious irritations, a plan had been devised by which thin silk strands were threaded through the flesh between the third and fourth metacarpal bones. Previous and later tests showed that no soreness or injury resulted from long attachment or from hundreds of miles of flying with the plates remaining in position.

The record of their returns as supplied by Mr. Lowell Gable is shown in Table I.

The magnets on the wings of the birds were separated about 12″ between centers when in a horizontal, full-spread position. They were of such strength as to produce a horizontal intensity, half-way between them, about equal to that of the earth's magnetic horizontal component in the region of State College. This value is approximately .172 oersted.

TABLE II.

3 with copper plates deviated less than 10° from a direct line toward the home loft, sixty-five airline miles to the northeast.
4 with copper were less than 50° off from this line toward the north.
3 others with copper were less than 30° off from this line toward the east.
4 with magnets were about 45° off from this line.
2 with magnets were more than 90° from this line.
3 with magnets were approximately 130° off from this line.
1 with magnets flew in the exactly opposite direction from the home loft.

By assuming that the magnets move up and down with approximately simple harmonic motion and estimating the wing beats at 180 per second and their amplitude at ⅛ foot, a value of .12 microvolt per centimeter was computed for the maximum induced e.m.f. per centimeter on the bird's anatomy.

This value of .12 microvolt per centimeter, as can be seen, is of the same order of magnitude as the change of e.m.f. involved in flying eighty miles toward or away from the normal magnetic-vertical line previously considered at 45° latitude. Both its magnitude and its pulsating character, i.e., changing three-hundred and sixty times per minute, from zero to .12 microvolt per centimeter and back to zero, might conceivably serve to confuse the bird and upset his navigating prowess during its occurrence.

The following important points should be noted:

1. Eight of ten birds with copper plates returned home during the first two days.

2. Of those returning, having originally had two magnets each and subsequently losing either one or both, only one with one magnet got home before the fourth day.

3. Only one bird with two magnets got home at all and it arrived on the fourth day.

4. None of the other birds, at this writing, has ever returned.

The final results of the experiment were well presaged by the directions in which the birds left the release point. Each was carefully watched out of sight before another was released. Table II

Table III. Showing release points and final locations of birds released.

| Release | No. | Band | Place found | Date | Wind velocity, dir. to lofts, temp., storms, etc. |
|---|---|---|---|---|---|
| Release No. 1 | 1 | AU43 PA11891 | St. Paul | 6/30 | 20 mi./hr. E.N.E. S.W. |
| St. Paul, Neb., | 2 | WSC41 D2020 | Grand Island | 7/1 | 71°F |
| June 29, 1944, | 3 | IF431 5014 | Grand Island | 7/1 | 18 thunderstorms during day |
| distance 43 miles, | 4 | AU43 U5115 | Ulysses | 7/1 | |
| 14 birds, | 5 | IF43 K4994 | Callaway | 7/3 | |
| 7 reported | 6 | AU43 V50825 | Franklin | 7/4 | |
| | 7 | IF43 K5013 | Belgrade | 7/5 | |
| Release No. 2 | 1 | AU44 P48107 | Clay Center | 7/2 | 22 mi./hr. S.E. N.W. |
| Hastings, Neb. | 2 | WB 31 | Greeley | 7/3 | 80°F |
| June 30, 1944, | 3 | AU44 P8113 | Harrard | 7/3 | 16 thunderstorms |
| distance 42 miles | 4 | AU43 V50824 | Heartwell | 7/2 | |
| 13 birds, | 5 | AU43 PA11886 | Hastings | 9/11 | |
| 5 reported | | | | | |
| Release No. 3 | 1 | AU44 V35151 | Sargent | 7/3 | 20 mi./hr. S. S.E. |
| Broken Bow, Neb., | 2 | AU44 V35165 | Ericson | 7/3 | temperature? |
| July 1, 1944 | 3 | AU44 P48110 | Bassett | 7/3 | 3 thunderstorms |
| distance 50 mi. | 4 | AU44 P48320 | Mason City | 7/2 | |
| 20 birds, | 5 | AU44 V35137 | Ansley | 7/8 | |
| 10 reported | 6 | AU44 V35160 | Ansley | 7/8 | |
| | 7 | AU43 V58822 | Comstock | 7/2 | |
| | 8 | IF43 BER414 | Broken Bow | 7/6 | |
| | 9 | AU43 U5389 | Taylor | 7/3 | |
| | 10 | AU43 V5111 | 12 W Loup City | | |
| Release No. 4 | 1 | WB26 | Heartwell | 7/4 | 24 mi./hr. S.S.W. N.N.W. |
| Minden, Neb., | 2 | AU44 V35162 | Kenesaw | 7/5 | 93°F |
| July 2, 1944 | 3 | AU44 P48117 | Longpine | 7/7 | 7 thunderstorms |
| distance 25 mi., | 4 | *AU43 Y4794 | 5 mi. E. of Kearney | 7/5 | *AU43 Y4794—a Heitzman bird |
| 15 birds, | 5 | AU43 X27252 | Kenesaw | 7/5 | came within 3 miles of finding |
| 7 reported | 6 | AU43 PA11780 | Holredge | 7/6 | loft |
| | 7 | AU43 M4820 | Lexington | 7/7 | |
| Release No. 5, | 1 | *WB 28 | East Loft | 7/5 | 21 mi./hr. S.E. |
| Havens, Neb., | 2 | AU44 26649 | Minden | 7/4 | 91°F |
| July 3, 1944 | 3 | AU44 V35141 | Lexington | 7/4 | 23 storms |
| distance 70 mi., | 4 | AU43 PA11696 | Ashton | 11/23 | *A Heissler (York, Pa.) bird flew 70 |
| 17 birds, | | | | | mi. and found loft. Direction was |
| 4 reported | | | | | away from State College |
| Release No. 6 | 1 | AU44 14356 | Gothenburg | 7/5 | 17 mi./hr. S.E. |
| Cozad, Neb., | 2 | AU44 H48122 | Burwell | 7/6 | 81°F |
| July 4, 1944, | 3 | AU44 B335150 | Cozad | 7/6 | 11 storms |
| distance 48 mi., | 4 | AU44 V35159 | Cozad | 7/6 | |
| 16 birds, | 5 | AU44 P48118 | 5 mi. N. Cozad | 9/1 | |
| 9 reported | 6 | WB 27 | Oconto | 9/15 | |
| | 7 | AU43 DC1486 | Dunning | 7/6 | |
| | 8 | IF43 B518 | Bertrand | 7/26 | |
| | 9 | AU43 J6506 | 11 mi. NW Miller | 9/5 | |

indicates the general direction at which the birds went out of sight from the release point.

Since these birds were well trained and outstanding homing pigeons, Mr. Gable and others expert in the field of flying and racing homing pigeons expected at least two or possibly three of those with two magnets to home, just on the basis of their intelligence and ability to "hunt" their way home. On the basis of probability and chance a similar prediction would obtain.

## EXPERIMENT II

## The First Nebraska Experiment

With the positive though limited data for the electromagnetic phase of the theory secured, the groundwork was laid for testing the theory as a whole. Examination of the U. S. Coast and Geodetic Survey Map I revealed that the conjugate point of State College, Pennsylvania, (corrected for 1944) i.e., another position having

TABLE III.—*Continued.*

| Release | No. | Band | Place found | Date | Wind velocity, dir. to lofts, temp., storms, etc. |
|---|---|---|---|---|---|
| Release No. 7<br>Columbus, Neb.,<br>July 7, 1944,<br>distance 100 miles,<br>2 birds released,<br>1 reported | 1 | AU44 P48109 | Thedford | 7/10 | 21 mi./hr. 94° 25 storms |
| Release No. 8<br>20 mi. east<br>Council Bluffs, Iowa,<br>July 8, 1944,<br>distance 185 mi.,<br>1 bird released,<br>1 received in west loft | 1 | IF43 B1014 | West loft | 7/10 | 21 mi./hr. N.E. 77°F 18 storms—this bird flew 185 miles in a direction opposite from State College and found the west loft near Kearney |
| Special Release No. 1<br>Lincoln, Neb.,<br>May 14, 1944,<br>distance 120 mi.,<br>12 birds,<br>5 reported | 1 | RBC5951JF | 15 mi. W. Grand Island | 5/16 | Day of release was clear, no winds, cool (Birds trained up to 50 and more miles) |
| | 2 | NPAB 3542 | Lenora, Kan. | 8/15 | |
| | 3 | AU-A 13726 | Fort Morgan, Col. | 5/24 | |
| | 4 | VE 14576 | Odell, Neb. | 5/22 | |
| | 5 | AU-Y-6005 | Freemont, Neb. | 5/16 | |
| Special Release No. 2<br>Lincoln, Neb.,<br>June 25, 1944,<br>12 birds released,<br>6 reported | 1 | E11712-43 | Lincoln, Neb. | 7/25 | |
| | 2 | 42W18225 | Lincoln, Neb. | 7/17 | |
| | 3 | NPA43 B13239 | Lincoln, Neb. | 7/25 | |
| | 4 | NPA44 B8117 | Lincoln, Neb. | 6/24 | Weather clear |
| | 5 | 41Z17857 | Wahoo, Neb. later Lindwood, Neb. | 6/26<br>7/6 | Slight south wind |
| | 6 | NPA43-B13241 | Hebron, Neb. | 6/30 | |
| Special Release No. 3<br>Berkeley Springs, Va.,<br>Cozad, Neb. birds<br>September 16, 1944<br>distance 100 miles<br>J. A. Rendle birds,<br>12 birds released,<br>6 reported | 1 | AU44 W8831 | Pittsburgh, Pa. | 9/20 | |
| | 2 | AU44 W8849 | Palmerton, Pa. | 10/2 | Weather was clear |
| | 3 | AU44 W8852 | Freeland, Pa. | 9/25 | Very light breeze |
| | 4 | AU44 W8841 | Near Berkeley Springs, Va. | 9/19 | Haze in distance |
| | 5 | AU43 A1436 | E. Liverpool, Ohio | 9/20 | Birds released in sets of 3's |
| | 6 | AU44 W3851 | Portage, Pa. | 9/26 | |

identical values of magnetic vertical intensity and Coriolis force, lies in about the south-central portion of Nebraska (ten miles north of Kearney). A plan was worked out whereby a group of one hundred or more carefully selected homing pigeons would be settled[10] and trained at State College, after which the birds and lofts would be transferred to the above conjugate point. Tests would then be made to determine if they would fly from the surrounding regions toward that point instead of heading for State College, Pennsylvania.

With this plan in view, training flights up to fifty miles around State College were decided upon and the birds were housed in a conspicuously designed and painted loft so it could be seen from as great a distance as possible. In addition, it was moved[11] each day so that the occupants, returning from training flights, would become accustomed to looking for the loft and not for familiar landmarks nearby. A five-foot observation balloon was raised each day about one-hundred and fifty feet above the loft as an additional beacon or guide. All these plans were executed carefully. The loft had a peak resembling a church spire and the whole was painted a brilliant yellow. It was moved two hundred feet each day and during the training period traveled a distance, by tractor power, of about three miles in a zigzag path over and across the college farms. The birds were given initial

---

[10] Having been in full flight around loft, the bird returns and enters it. During his first free flight outdoors even though raised on the spot, it is easily frightened off in a panic, and may fail to take visual account of its surroundings. On a second release it is practically certain to return.

[11] Its location was shifted 200 feet each day for training flights of the birds, and never returned to a previous location.

piloting training flights[12] of one, two, four, eight, and sixteen miles in several directions around the compass. From then on for a period of three months they had practice flights twice daily up to fifty and more miles. There was seldom a half-day during which they were not released at some new point around the compass. Care was used to prevent repetitional flights which might either aid or hinder them in the Nebraska experiment. It is known that repeated training during the same part of the day and over the same terrain causes homing pigeons to fly habitually by the sun's direction. Near the end of the training period they were released from time to time in singles, doubles, and sets of five at twenty-five-mile airline distances.

On the day set for the start of the trip to Nebraska the birds were transported to Altoona by truck in one of the homing lofts. The shipment from there on was by express so the loft had been carefully designed to make it as large as possible and yet have it go inside the express car. During the entire 1100-mile trip the nests with eggs and young were cared for by the parent birds. On arrival in Kearney the loft was set on the fair grounds until magnetic measurements were made to check the values obtained from the U. S. Department of Commerce. Also a second loft had been built and previously shipped, to double the chance of the birds' finding one in case they navigated from the release points to the predicted conjugate point. Askania magnetometer measurements showed that the magnetic vertical-field intensity ten miles north of Kearney was the same as State College. Unfortunately, however, the field intensity gradually increased to about three times its normal value fifteen miles east and returned to its normal value ten miles beyond the anomaly. This condition meant that there were actually two conjugate points twenty-five miles apart, which called for a choice of placing the two lofts about 10 miles apart on either side of the point as planned according to the U. S. averages, or putting one at the normal conjugate point and the other twenty-five miles to the east at the anomalous conjugate point. The limited personnel and transportation facilities made the twenty-five mile separation impractical. In addition, magnetic anomalies exert most of their influence fairly near the ground surface and for all these reasons the former strategy was used. The lofts were carefully placed five miles east and west of the calculated position on as high ground as could be found in that flat country. The birds were allowed to rest one day to acclimate themselves to the immediate surroundings.

Beginning with June 28, groups of birds were taken out to release points around the compass forty and more airline miles away. Each bird carried an army-message capsule with the words.

Important

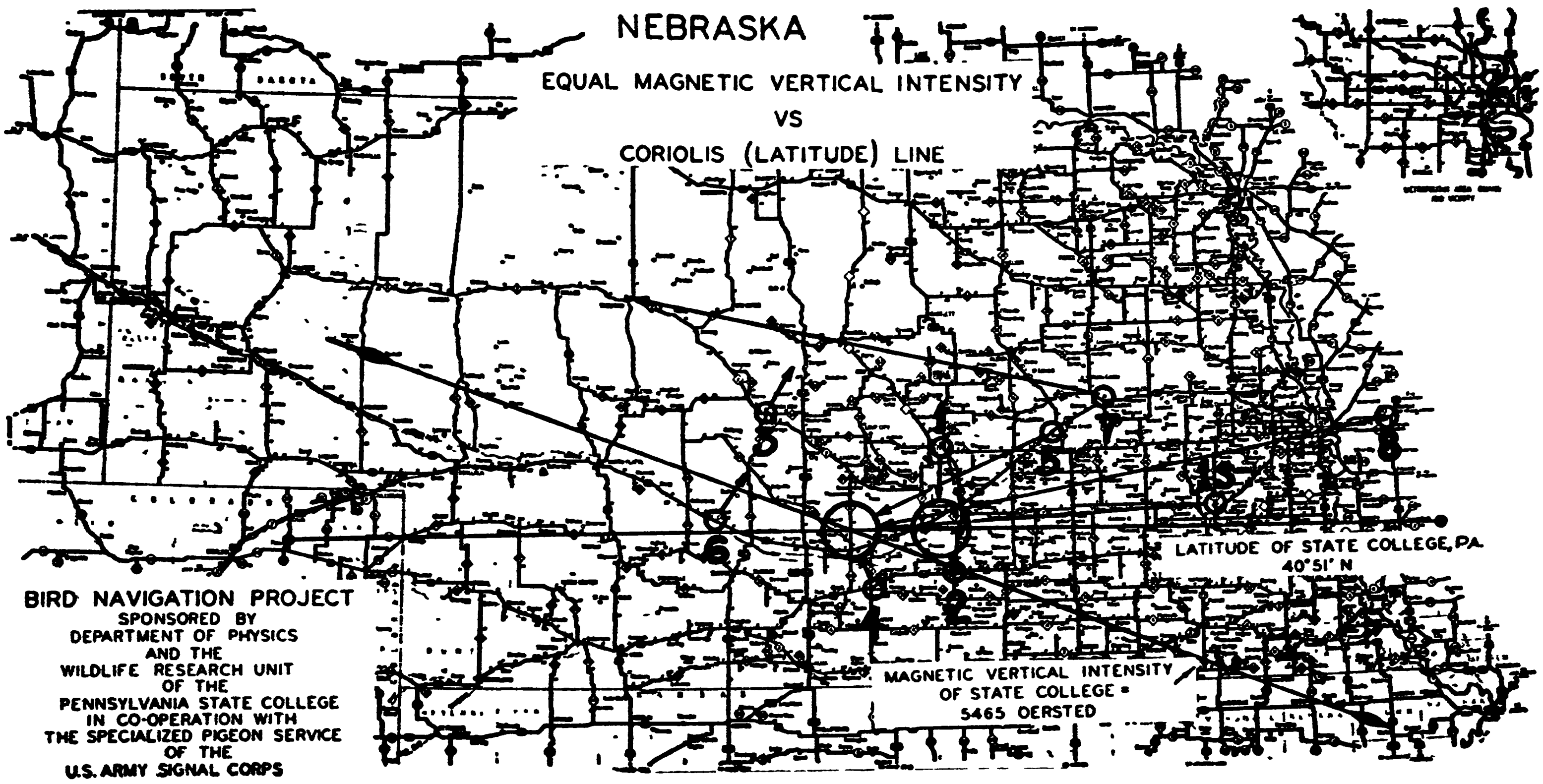
Telegraph Immediately (collect)
Wildlife Research Unit
cooperating with
The Army Pigeon Service
c/o Fort Kearney Hotel
Kearney, Nebraska
Give band number
Place found
Date

The data obtained from these experiments are set forth in Table III. Releases No. 1 through No. 8 show the direct result of the experiment proper carried out between June 29 and July 8. Releases labeled Specials No. 1 and No. 2 were performed before, and Special No. 3 after the main experiment and will be discussed later in the report. Map III, showing the state of Nebraska, contains a representation of the data shown on Table III except for that designated Special release No. 2 and No. 3. Each vector on the map represents graphically the sum total of the results for a given day or release number, the latter being in one-to-one correspondence on the map and table.

The flights of all birds reported were plotted as vectors on a preliminary work map and their vector sum obtained. The length of the latter was divided by the number of flights involved and the resultant vector represents the average result of all the recorded flights of that day. In referring to these in this report the term "total flight vector" will be used. A brief inspection of the total flight vectors for the eight days shows that numbers one, two, four, five, and eight not only point in the right direction but actually

---

[12] Short distance flights in which guidance could always be had by means of familiar landmarks.



MAP III. Flight vectors showing average flights of the various groups of pigeons released according to data on Table III.

terminate within ten miles of one or the other of the two predicted points (the normal and anomalous ones). In fact, four, five, and eight, and Special No. 1 terminate with good precision. It is a notable fact that one bird, a blue bar female, WB28, of flight number five, flew the seventy miles, and found and entered the loft, despite hot, twenty-seven-mile per hour winds at right angles to its line of flight. Another unusually significant flight was that of the light blue check female, band number IF43B1014, the only bird flown that day, and the only one released outside of Nebraska. Its flight started at 3:30 P.M. about twenty miles east of Council Bluffs, Iowa. Not only did it fly the one hundred eighty-five airline miles across hot winds averaging twenty miles per hour but, as with the seventy-mile flight described above, it flew in the opposite direction from its home in State College to find and enter its 4×4×5 foot loft displaced one thousand miles west of its home to the middle of a Nebraska grain field.

Total flight No. 6, representing nine reported birds, shows one of two anomalous results; these birds flew in an average direction 60° off the theoretical course. One possible explanation is that there was a violent thunder storm between the release and loft positions that come up swiftly just after the last bird had been released. An important fact about this flight (true for all releases), is that the birds did not fly toward State College nor in random directions. None of the nine flew in a direction as much as 90° off from the direction of the total flight vector. It is interesting to observe that, although the trend of this flight was not toward the home Coriolis line, it did follow almost the shortest route to the correct magnetic-vertical line of State College. Flight No. 3, also anomalous, has not the mitigating factor of an intervening electric storm, but it should be noted that the magnetic fields in the region involved are of the most distorted on record, according to Mr. James Affleck of The Gulf Research and Development Company. Release No. 7 is relatively unimportant since only one bird was reported of the two released. Actually, the flight direction is quite good, being only about 40° off theoretical, and the bird may have flown in and out of the twenty-five-mile error distance for homing pigeons. Further, the hot Nebraska wind that day varied from twenty-six to forty-eight miles per hour along the flight path, and the temperature averaged 94°F throughout.

Special release No. 1 represents an "advance guard" experiment in which twelve carefully trained homing pigeons were shipped to Lincoln, Nebraska, to be released by Professor P. A. Downs of the University of Nebraska. These birds, although not trained specifically for the experiment, had had several years of navigational and racing experience from points up to one hundred miles from State College, Pennsylvania. After they had been singly released and carefully watched out of sight, Dr. Downs later remarked that although they looked like good homers, unfortunately they went the wrong direction. Eight of the twelve had flown out of sight toward the west, exactly opposite from State College. Examination of the total flight vector for this group reveals that, for the five reported birds, making an average flight of one-hundred thirty miles, it ends practically on the predicted position.

On June 25, Dr. Downs again released twelve homing pigeons of excellent flying stock raised in State College, Pennsylvania, this time "prisoners" up to a year or more in age, and without flying experience except inside their home loft and fly-pen. In observing their behavior at the liberation, it was noted that they merely scattered in random directions. Of six subsequently reported by telegraph, four remained at Lincoln, two others went in directions neither toward State College nor its Nebraska conjugate. This serves to illustrate the fact that even good homing pigeons must have flying experience and training, in order to respond to the navigating influences.

Special release No. 3 was the outcome of a visit to Mr. J. A. Rendle of Cozad, Nebraska, an outstanding racing pigeon expert, whose name had been given us by Major Otto Meyer of the U. S. Army Signal Corps. Informed of our experiment, he straightway offered to train some of his best young birds for a reverse test of the theory. The conjugate point of Cozad is somewhat east of the middle of Pennsylvania, but the magnetic deposits in the mountains scatter it into several positions east and west along the

Coriolis line which is essentially the same (.04 of a degree north) as that of State College, Pennsylvania.

On September 15, we received from Mr. Rendle twelve young homing pigeons trained around the compass fifty miles from Cozad. The following day they were released in groups of three, at Berkeley Springs, Virginia. Inspection of Pennsylvania Map IV shows the total flight vector for the six reported: one stayed at or returned to Berkeley Springs but none followed a flight line more than 60° or so from the total flight vector, itself in a correct direction, and terminating within ten miles of the same Coriolis line through Cozad. This Cozad, Nebraska, to Pennsylvania result is a corroboration of the strongly indicative State College, Pennsylvania, Kearney, Nebraska, results.

Shortly before receiving these young Cozad birds Mr. Rendle shipped 10 old birds which had had 200 to 400 mile flights in their records. These were released at various places one hundred miles or more from the center of Pennsylvania. Closely paralleling Mr. Rendle's own predictions three of these birds homed across the 1400 or more miles to Cozad in from one to three weeks. Finally three of the remaining ones were reported found in the region of Michigan, western Pennsylvania, and eastern Ohio. A full discussion of these results requires more space than is available here and will appear in a later paper.

### *Salient Points of the First Nebraska Experimental Results*

1. Of one-hundred twenty-two birds trained in State College and released in Nebraska, only one was ever reported east of Indiana. This was telegraphed in eight months later from Phillipsburg, Pennsylvania.

2. Of the eight main Nebraska releases, six of the total flight vectors support the validity of the theory, in addition to the indicative results obtained with the Cozad, Nebraska birds flown from south of Pennsylvania.

3. The Askania magnetometer readings on a sixty-mile line indicate there are two rather than one magnetic vertical-Coriolis conjugates in the Kearney, Nebraska, region. Two groups of total flight vectors (Special No. 1 together with Nos. 4, 5, and 8) and (Nos. 1 and 2), also show this to be true.

4. Two birds, one from seventy, the other from one-hundred eighty miles, flew in directions almost exactly opposite to their home area in State College, Pennsylvania, and found their lofts at the predicted conjugate point. An additional bird was found within three miles of the west conjugate and three were reported within less than ten miles of the anomalous conjugate point. Considering the fact that a homing pigeon should be familiar with landmarks twenty-five miles out from its home loft in order to assure a successful flight termination, it is evident that six of the reported birds (see data on releases Nos. 1-2 and 3, 4–4, 5-1, 8–1, Special No. 1–1) performed the equivalent of "homing" at their conjugate points. It is probable that a high percent of those used in the experiment flew into the regions involved, but, not seeing the loft or familiar landmarks, flew out again in random directions.

5. The results were good in the face of many adverse factors acting against success, such as:

(a) The birds were trained in green mountainous country but the experimental flights were over flat, brown terrain.

(b) The limited supply of five-foot signal balloons used in piloting the birds at the home loft were quickly blown down and punctured on stones during the first few minutes in the air during the first two days in Nebraska so their use had to be abandoned.

(c) The experimental flights were executed in hot, dry winds of from sixteen to forty-eight miles per hour velocity which blew continuously during the day. Experienced racing-pigeon men fear any wind velocities above twelve miles per hour even at normal temperatures. This was in direct contrast to low velocity winds or complete absence of winds during the training period in Pennsylvania.

(d) The birds of necessity were always liberated after 1:00 P.M., because of inadequate help and transportation. This is late in the day for the most successful flying and navigating results.

(e) Experience has shown the longer experimental flight to be more successful. Over-caution, especially at the beginning of the undertaking,

TABLE IV.

| No. bird | Band No. | Year | Color | | Age (months) |
|---|---|---|---|---|---|
| 1 | US44SC 4950 | 44 | B.Ch. | 4 | B–Camp Crowder, M–Ft. Monmouth, S–Lebanon, RM–Ft. Monmouth |
| 2 | US44SC 20400 | 44 | Dk.Ch. | — | B–Camp Edison, N. J., M–Ft. Monmouth |
| 3 | US44SC 25045 | 44 | Griz. | 7 | B–Ft. Sam Houston, S & RN–W. Va., RS & RM–Ft. Monmouth |
| 4 | US44SC 24545 | 44 | Red | 6 | B–Ft. Sam Houston, M–Ft. Monmouth, S, C, RM–Lebanon |
| 5 | US44SC 21496 | 44 | B.Ch. | — | B–Camp Clairborne, M–Ft. Monmouth |
| 6 | US44SC 142 | 43 | B.Ch. | — | B–Indio, Cal., S & F–Ft. Monmouth, S & M–Lebanon, RS & F–Ft. Monmouth |
| 7 | US44SC 22935 | 43 | B.Ch. | 10 | B–Ft. Benning, S & M–Ft. Monmouth, RS & M–Lebanon, B–Camp Crowder |
| 8 | US44SC 21480 | 44 | Blue | — | B–Camp Clairborne, M–Ft. Monmouth |
| 9 | US44SC 12255 | 43 | B.Ch. | 19 | B–Camp Crowder, S & M–Ft. Monmouth, RS & RM–Lebanon |
| 10 | US44SC 20241 | 44 | Red | — | B–Camp Clairborne, S & M–W. Va., RS & RM–Ft. Monmouth |
| 11 | US44SC 4369 | 44 | B.Ch. | 5 | B–Camp Crowder, M–Ft. Monmouth, S–Lebanon |
| 12 | US44SC 25569 | 44 | B.Ch.Spl. | 5 | B–Ft. Sam Houston, S & M–Lebanon, RS & RM–W. Va., RM–Ft. Monmouth |
| 13 | US44SC 37887 | 43 | Blue | — | B–Baltimore, Md., S & F–Ft. Monmouth |
| 14 | US44SC 21946 | 44 | Dk.Ch. | — | B–Camp Clairborne, M–Ft. Monmouth |
| 15 | US44SC 25520 | 44 | Red | 6 | B–Ft. Sam Houston, S & M–Lebanon, RS & RM–W. Va., RM–Ft. Monmouth |
| 16 | US44SC 4904 | 44 | Dk.Ch. | 4 | B–Camp Crowder, S & M–Lebanon, RM–Ft. Monmouth |
| 17 | US44SC 21479 | 44 | Blue | — | B–Camp Clairborne, M–Ft. Monmouth |
| 18 | US44SC 21487 | 44 | B.Ch. | — | B–Camp Clairborne, M–Ft. Monmouth |
| 19 | US44SC 24033 | 44 | B.Ch. | 8 | B–Ft. Sam Houston, S & M–Lebanon, RS & RM–W. Va., RS & RM–Ft Monmouth |
| 20 | US44SC 34217 | 43 | Dk.Ch. | — | B–Elizabeth, N. J., M–Ft. Monmouth, RM–Lebanon |
| 21 | US44SC 21477 | 44 | B.Ch.W.F. | — | B–Camp Clairborne, M Ft. Monmouth |
| 22 | US44SC 21288 | — | — | — | — — — — |
| 23 | US44SC 24084 | 44 | B.Ch. | 7 | B–Ft. Sam Houston, S & M–Lebanon, RS, RM, C–Ft. Monmouth |
| 24 | US44SC 21178 | 44 | Red | — | B–Camp Clairborne, M–Ft. Monmouth |
| 25 | US44SC 41183 | 44 | B.Ch. | — | B–Camp Clairborne, S & M–Ft. Monmouth |
| 26 | US44SC 21465 | 44 | Red | — | B–Camp Crowder, M–Ft. Monmouth |
| 27 | US44SC 5159 | 44 | B Ch. | 4 | B–Camp Crowder, M–Ft. Monmouth |
| 28 | US44SC 25208 | 44 | Blue | 6 | B–Ft. Sam Houston, M–Ft. Monmouth, S & RM–Lebanon |
| 29 | US44SC 21950 | 44 | Red | — | B–Camp Clairborne, M–Ft. Monmouth |
| 30 | US44SC 12134 | 43 | Dk.Ch.W.F.Bl. | 11 | B–Camp Crowder, S & M–Ft. Monmouth, RS & RM–Lebanon |
| 31 | US44SC 40128 | 44 | Blue Tk. | — | B, S & F–Camp Clairborne, RS & F–Ft. Monmouth |
| 32 | US44SC 22063 | 43 | B.Ch.W.F. | 11 | B–Ft. Benning, S & M–Lebanon, RS & RM–W. Va., RS & RM–Ft. Monmouth |
| 33 | US44SC 5403 | 44 | B.Ch.Spl. | 4 | B–Camp Crowder, S & M–Ft. Monmouth, RM–Lebanon |
| 34 | US44SC 1732 | 44 | Blue | 7 | B–Camp Crowder, S & M–Lebanon |
| 35 | US44SC 46539 | 43 | B.Ch. | — | B–Camp Clairborne, S & F–Hackettstown, N. J., RS & F–Ft. Monmouth |
| 36 | US44SC 2499 | 44 | Red | 7 | B, S & F–Camp Crowder, RS & M–Lebanon |
| 37 | US44SC 4184 | 44 | Mosaic | 5 | B–Camp Crowder, M–Lebanon |
| 38 | US44SC 41344 | 43 | Red | — | B–Camp Clairborne, S & F–Ft. Monmouth |
| 39 | US44SC 41043 | 43 | Blk.Ch. | — | B–Camp Clairborne, S & M–Ft. Monmouth |
| 40 | US44SC 40246 | 44 | B.Ch. | — | B, S & F–Camp Crowder, RS & F–Ft. Monmouth |
| 41 | US44SC 1595 | 44 | B.Ch. | 7 | B–Camp Crowder, S & M–Lebanon, RM–Ft. Monmouth |
| 42 | US44SC 22460 | 43 | Blue | 10 | B–Ft. Benning, S & M–Lebanon |
| 43 | US44SC 23213 | 43 | Blue Ch. | — | B–Ft. Storey, Va., S & M–Ft. Monmouth |
| 44 | US44SC 40990 | 43 | Blue Tk. | — | B–Camp Clairborne, S & F–Ft. Meade, Md., S & F–Ft. Monmouth, RS & M–Lebanon |
| 45 | US44SC 24705 | 44 | Silver | — | M–Ft. Monmouth, S–W. Va. |
| 46 | US44SC 20241 | 44 | B.C. | — | B–Camp Clairborne, M–Ft. Monmouth |
| 47 | US44SC 22209 | 43 | B.Ch. | — | B–Ft. Benning, S & M–Lebanon |
| 48 | US44SC 22191 | 43 | Blk. | 11 | B, S & F–Ft. Benning, RS & M–Lebanon, RM–Ft. Monmouth |
| 49 | US44SC 44277 | 44 | Blue | — | B, S & F–Camp Clairborne, RS & F–Ft. Meade, RS & F–Ft. Monmouth |
| 50 | US44SC 25116 | 44 | B.Ch.Tk. | — | S & M–Lebanon, RS & RM–W. Va., RM–Ft. Monmouth |

### Glossary to Table IV

Bred (B)—includes mating of old birds, hatching and raising youngsters to flying stage (about four and one-half weeks or longer), but without any flying outside of the loft confines.

Settled (S)—involves letting birds fly outside of loft for the first time. Youngsters up to five or six weeks of age can be settled at a new loft, any distance whatever from the home cote, twenty-four hours after arrival.

Resettled (RS)—repeating operation (S) at a new loft or loft location. The Army was able to do this in seven or eight days by mobiling the birds at the initial settling station.

Flown (F)—this means that, in addition to regular daily exercise flights from the loft, the birds are each day released at progressively greater distances from the home location. Flights of under twenty-five miles are for the purpose of giving them knowledge of local landmarks for piloting to the exact loft locations, plus knowledge of magnitude and gradients of the directives for later use in navigating over unfamiliar territory.

Mobiled (M)—after settling to a portable loft, the latter is moved each day so birds, in their local flying and short training flights, search for it rather than for a particular spot on the land. In military mobiling operations the loft transfers are along a straight line. In the training operations of these experiments the successive locations were always on a circuitous route.

Remobiled (RM)—after settling and mobiling in a given region the birds are resettled, with the same or similar appearing portable lofts, at any desired distance from the last location. In the war, birds were mobiled in this country and remobiled in the European and Pacific war theatres.

caused us to keep the release distances on the short side.

At the completion of the first tests of the bird navigation theory just described, the data obtained were examined critically by the statistical group at the Iowa State College. These men, under Dr. George W. Snedecor, recommended that, in view of the indicative results, work should be continued and more data should be obtained. With this object in view plans with newly suggested procedures were carried out which resulted in the following experiments.

### EXPERIMENT III

### Hamlet Experiment, September 1944

The writer has talked to many persons, with much experience in racing homing pigeons, who contend that birds have a mysterious power to retrace a path over which they have been carried, even in darkness or in a drugged condition. Some tests of this contention[13] have been made and negative results obtained. There are also many experienced pigeon-racing men who claim that homing pigeons possess the remarkable ability to return to a loft from a distance, merely by virtue of their having lived in it. In the experiment of June, 1944, already recorded, the birds used in "Special release No. 2," had neither flying experience around their home loft at State College, Pennsylvania, nor navigational training. Although the data are far too limited to be conclusive, they do indicate an inability of such birds to make any effective attempt whatever at homing.

At the close of the first full-scale experiment in Nebraska it was found possible to secure, from mobile lofts of the Army Signal Corps at Fort Monmouth, New Jersey, three hundred pigeons, which (with one exception, as later learned) had had no training in navigation, although some had had limited "pilotage" experience from being settled and flown from short distances at the lofts where they were bred. These birds were bred and shipped from Fort Sam Houston, Texas, Camp Crowder, Missouri, and other stations indicated in Table IV and had been given only "mobile" training at Fort Monmouth. Upon arrival at the experimental headquarters at Hamlet, Nebraska, on September 29, 1944, they were placed in Army mobile lofts and permitted a twenty-four-hour rest before being transferred to the various release points. In order to take care of the slight possibility that they might be able to navigate even with the various types of training indicated, the releases were arranged from locations within a hundred and fifty miles radius of Hamlet, Nebraska. The latter is practically on the conjugate point for Fort Monmouth, New Jersey, and according to the theory and the Kearney, Nebraska, results, the point toward which the birds would fly, should their limited Fort Monmouth mobile training be adequate. On September 30, 1944, two hundred and fifty of these birds were re-

---

[13] G. Reynaud, "The laws of orientation among animals," Smith. Inst. Ann. Report of Board of Directors, p. 481, (translated from Rev. des Deux Mondes 146) 380 (1898), also O. H. Mowrer, J. compar. Psychol. 19, 177 (1935).

TABLE V.

| Bird No. | Date | Place released | Place found | Distance traveled |
|---|---|---|---|---|
| | | | Group No. 1 | |
| 1 | Sept. 30, 1944 | Ansley | Ansley | 0 miles |
| 2 | Sept. 30, 1944 | 8 mi. E. Goodland | 5½ mi. E. Selden | 60 |
| 3 | Sept. 30, 1944 | Between Quinter and Ellis | Between Quinter and Ellis | 0 |
| 4 | Sept. 30, 1944 | 9 mi. W. Burlington | 5 mi. N.E. Kit Carson | 40 |
| 5 | Sept. 30, 1944 | 17 mi. E. Goodland | 17 mi. S.W. Burlington | 55 |
| 6 | Sept. 30, 1944 | Miller | Miller | 0 |
| 7 | Sept. 30, 1944 | 13 mi. S. Alma | Alton, Kan. | 40 |
| 8 | Sept. 30, 1944 | 3 mi. W. Bur. Stratton | 3 mi. W. Bur. Stratton | 0 |
| 9 | Sept. 30, 1944 | Phillipsburg, Kan. | Stocton S.E. | 23 |
| 10 | Sept. 30, 1944 | Ansley | Ansley | 0 |
| 11 | Sept. 30, 1944 | Miller | Holstein | 50 |
| 12 | Sept. 30, 1944 | 21 mi. E. Quinter | 7 mi. S.W. Waukaneg, Kan. | 7 |
| 13 | Sept. 30, 1944 | 3 mi. E. Goodland | Goodland | 3 |
| 14 | Sept. 30, 1944 | 5 mi. E. Goodland | Winona | 30 |
| 15 | Sept. 30, 1944 | 6 mi. N. Ansley | 4 mi. S.E. Broken Bow | 11 |
| 16 | Sept. 30, 1944 | 6 mi. E. Quinter | 20 mi. S.E. Grainfield | 17 |
| 17 | Sept. 30, 1944 | Goodland, Kan. | Goodland, Kan. | 0 |
| 18 | Sept. 30, 1944 | 12 mi. N. Ansley | 4 mi. N. Wood River | 58 |
| 19 | Sept. 30, 1944 | Quinter | Quinter | 0 |
| 20 | Sept. 30, 1944 | 23 mi. N. Miller | York, Neb. | 100 |
| 21 | Sept. 30, 1944 | 3 mi. N. Phillipsburg | 4 mi. N. Phillipsburg | 1 |
| 22 | Sept. 30, 1944 | Phillipsburg | Phillipsburg | 0 |
| 24 | Sept. 30, 1944 | 5 mi. N. Miller | Juniata | 50 |
| 25 | Sept. 30, 1944 | Goodland | Goodland | 0 |
| 26 | Sept. 30, 1944 | 17 mi. E. Goodland | 8 mi. S. Monument | 32 |
| 27 | Sept. 30, 1944 | 14 mi. S. Miller | Elm Creek | 30 |
| 28 | Sept. 30, 1944 | 3 mi. W. Burlington | W. Eckley, Col. | 57 |
| 29 | Sept. 30, 1944 | 20 mi. E. Goodland | 19 mi. E. Goodland | 1 |
| 30 | Sept. 30, 1944 | Alma | E. Alma | 1 |
| 32 | Sept. 30, 1944 | Ellis, Kan. | Ellis, Kan. | 0 |
| 33 | Sept. 30, 1944 | 18 mi. N. Ansley | 8 mi. N.E. Broken Bow | 7 |
| 34 | Sept. 30, 1944 | Burlington | Burlington | 0 |
| 35 | Sept. 30, 1944 | Ogalla | Ogalla | 0 |
| 37 | Sept. 30, 1944 | 6 mi. E. Quinter | 11 mi. S.W. Stockton | 47 |
| 38 | Sept. 30, 1944 | 18 mi. E. Waukeeny | Bazine, Kan. | 40 |
| 39 | Sept. 30, 1944 | 3 mi. S. Alma | Republican City | 0 |
| 41 | Sept. 30, 1944 | 12 mi. E. Quinter | 5 mi. N.E. McCracken | 40 |
| 42 | Sept. 30, 1944 | Waukeeny | Utica | 30 |
| 43 | Sept. 30, 1944 | 11 mi. N. Miller | 1 mi. W. Heartwell | 48 |
| 44 | Sept. 30, 1944 | 6 mi. W. Burlington | Stratton, Col. | 19 |
| 45 | Sept. 30, 1944 | 3 mi. W. Burlington | 10 mi. W. Burlington | 7 |
| 46 | Sept. 30, 1944 | Ansley | Ansley | 0 |
| 47 | Sept. 30, 1944 | Ellis, Kan. | Ellis, Kan. | 0 |
| 50 | Sept. 30, 1944 | Alma | Alma | 0 |
| | | | Group No. 2 | |
| 23 | Sept. 30, 1944 | Goodland, Kan. | 7 mi. E. Wallace, Neb. | 110 (navigating experience at Ft. Monmouth. Flew toward conjugate point) |
| | | | Group No. 3 | |
| 31 | Sept. 30, 1944 | 9 mi. W. Burlington | 30 mi. N.E. Garden City, Kan. | 120 |
| 36 | Sept. 30, 1944 | 12 mi. S. Alma | 16 mi. N. Great Bend, Kan. | 115 (made long flights in general direction of place bred) |
| 40 | Sept. 30, 1944 | 15 mi. E. Quinter | Sharon | 150 |
| 48 | Sept. 30, 1944 | 9 mi. N. Phillipsburg | 3 mi. W. Rozel | 110 |
| 49 | Sept. 30, 1944 | Phillipsburg, Kan. | Haviland, Kan. | 145 |

(Towns in Nebraska unless otherwise specified)

leased in four groups near Goodland, Kansas, Miller, Nebraska, Alma, Nebraska, and Quinter, Kansas, respectively. Small groups of twos, threes, fours, and fives were released, ten minutes apart, beginning at dawn. The weather was cool and clear with gusty winds of about fifteen miles per hour blowing during the day. Of two hundred and fifty birds released, fifty-four were reported on or before October 30, by telegrams.[14]

---

[14] Of the remaining fifty of the original three hundred, half were youngsters, the others yearlings. The twenty-five yearlings were released at Culbertson in a group most of

The performance of the birds released, as indicated by the telegrams reporting the flight terminals, seems to divide them into three natural groups (Table V) each of which will be dealt with separately.

### *Results*

*Group No. 1.* These forty-eight birds flew in short random flights ranging from zero to sixty miles from the points of release. The arithmetical average of the forty-eight flights is about twenty miles. Of these, twenty were reported within a mile of the release point. An even more significant indication about the flight of these birds is that the average of the vector sum of their flight vectors (previously referred to as the "total flight vector") is less than ten miles. Since these flights, including distance and directions, are vector quantities, the vector sum and its average is the only one which has real meaning in this connection.

*Group No. 2.* This 110-mile flight seems significant in that this bird flew in a direct line for Hamlet, Nebraska, which is the conjugate point of Fort Monmouth, New Jersey, and stopped just thirty miles beyond. Of the fifty-two birds reported, it is the only one which had had the proper training in navigating to do this. It was later learned that it had been used in communications at Fort Monmouth, New Jersey, and inadvertently was shipped with the less experienced birds.

*Group No. 3.* These five birds are distinguished from the rest in that they made long flights averaging one hundred and fifteen miles in length, in addition to flying in directions quite accurately toward their original homes which are as follows:

| Bird No. | Place bred |
|---|---|
| 31 | Camp Clairborne, Louisiana |
| 36 | Camp Crowder, Missouri |
| 41 | Camp Crowder, Missouri |
| 50 | Fort Benning, Georgia |
| 51 | Camp Clairborne, Louisiana |

A check made with the head of the Army Pigeon Service revealed that they had been settled and had had considerable flying experience around their home loft before being shipped to Fort Monmouth, New Jersey, which may account not only for the length but also the direction of their flights.

It should be noted that any one of the fifty reported birds, having been bred from some of the world's finest homing pigeons, had the inherited qualities and the mental and physical make-up to fly a like distance toward Fort Monmouth or its conjugate point had they also possessed the means of so doing, brought about by the required type of navigational training. If the "retracement" contention had credence, they should have flown back along the path over which they had been transported by train two days before, or retraced the paths over which they had been taken by car from Hamlet to the various release points, during the hours previous to their release. In the latter case they should have flown toward Hamlet first and then started eastward to Fort Monmouth instead of the short random flights in forty-seven of fifty-four reported cases.

It is a well established fact that homing pigeons settled and trained to navigate at any point, will, in most cases, if physically able, return to it when released moderate distances away, such as one to three hundred miles. This experiment indicates, however, that just settling[15] and mobiling will not replace the training in navigation which they must have to home successfully from a distance. It therefore seems evident, whereas they may inherit the ability to navigate, that they must first have specialized and adequate training in order to accomplish it.

The results of the Hamlet experiment and Special No. 2 and all other tests of these experiments, indicate that homing pigeons do not retrace paths over which they have just been transported unless these latter happen to be directly in line from the home (or its conjugate point) to which they have been trained to pilot and to navigate.

---

which took up residence on the spot. The twenty-five youngsters were retained for "navigation" training at State College, Pennsylvania, for later experimentation.

[15] See explanation in Glossary appended to Table IV. Explains training which they must have to home successfully from a distance. It therefore seems evident, whereas they may inherit the ability to navigate, that they must first have specialized and adequate training in order to accomplish it.

TABLE VI. Showing release and terminal points of birds reported or which returned to the home loft at State College, Pennsylvania. Releases by Major Otto Meyer, Lt. A. M. Lehman, Dr. P. F. English, Jack Nesbitt, Roy Johnson, Donald Brenner, H. L. Yeagley, Dr. George O. Hendrickson. (X)—Birds released within 100 miles of the line of tangency and crossed over. E—Birds released east of line of tangency. W—Birds released west of line of tangency.

| Yellow band number | U. S. Army Signal Corps band No. | Place of release | Date | Place found | Date | Designation |
|---|---|---|---|---|---|---|
| 201 | US45SC 68367 | Marshaltown, Iowa | 6/12 | 11 mi. E. of Sious Falls, S. D. | 6/14 | W |
| 202 | US45SC 68272 | Joliet, Ill. | 6/11 | La Salle, Ill. | 6/16 | (X) |
| 203 | US45SC 68390 | Arlington, Neb. | 6/12 | Nebraska City, Neb. | 6/14 | W |
| 204 | US45SC 68300 | Ames, Iowa | 6/12 | Tennanet, Iowa | 6/25 | W |
| 205 | US45SC 68345 | Morrison, Ill. | 6/14 | Marion, Ind. | 6/14 | (X) |
| 206 | US45SC 68343 | Cedar Rapids, Iowa | 6/12 | Casey, Iowa | 6/14 | W |
| 207 | US45SC 68352 | Carroll, Iowa | 6/12 | Not reported | | |
| 208 | US45SC 68350 | Arlington, Neb. | 6/12 | Not reported | | |
| 209 | US45SC 68355 | Morrison, Ill. | 6/11 | Gridley, Ill. | 7/3 | (X) |
| 210 | US45SC 68392 | Tiffin, Ohio | 6/11 | Not reported | | |
| 211 | US45SC 68368 | Wadsworth, Ohio | 6/10 | State College, Pa. | 6/20 | E |
| 212 | US45SC 68369 | Carroll, Iowa | 6/12 | Stuart, Iowa | 6/24 | W |
| 213 | US45SC 68356 | Carroll, Iowa | 6/12 | Not reported | | |
| 214 | US45SC 68380 | Dunlap, Iowa | 6/14 | Fontanelle, Neb. | 6/14 | W |
| 215 | US45SC 68389 | Altoona, Pa. | 6/10 | State College, Pa. | 6/10 | E |
| 216 | US45SC 68354 | Tiffin, Ohio | 6/11 | Hartville, Ohio | 6/14 | E |
| 217 | US45SC 68382 | Cedar Rapids, Iowa | 6/12 | Not reported | | |
| 218 | US45SC 68185 | Arlington, Neb. | 6/12 | Not reported | | |
| 219 | US45SC 68239 | Ottowa, Ohio | 6/11 | Not reported | | |
| 220 | US45SC 68351 | Arlington, Neb. | 6/12 | Not reported | | |
| 221 | US45SC 68344 | Valparizo, Ind. | 6/11 | Not reported | | |
| 222 | US45SC 68341 | Joliet, Ill. | 6/11 | Not reported | | |
| 223 | US45SC 68372 | Youngstown, Ohio | 6/10 | Loyalhanna, Pa. | 6/11 | E |
| 224 | US45SC 68353 | Carroll, Iowa | 6/12 | Coon Rapids, Iowa | 6/14 | W |
| 225 | US45SC 68364 | Van Wert, Ohio | 6/11 | Not reported | | |
| 226 | US45SC 68391 | Waterman, Ill. | 6/11 | Mt. Ayr, Iowa | 6/14 | W |
| 227 | US45SC 68362 | Butler, Pa. | 6/10 | State College, Pa. | 6/10 | E |
| 228 | US45SC 68381 | Youngstown, Ohio | 6/10 | State College, Pa. | 6/11 | E |
| 229 | US45SC 68385 | Indiana, Pa. | 6/10 | Johnstown, Pa. | 6/12 | E |
| 230 | US45SC 68363 | Ottowa, Ohio | 6/11 | Weirton, W. Va. | 6/20 | E |
| 231 | US45SC 68377 | Waterman, Ill. | 6/12 | Chariton, Iowa | 6/12 | W |
| 232 | US45SC 68273 | Arlington, Neb. | 6/12 | Garland, Neb. | 6/26 | W |
| 233 | US45SC 68394 | Valparizo, Ind. | 6/11 | Kankakee, Ill. | 6/13 | (X) |
| 234 | US45SC 68393 | Butler, Pa. | 6/10 | Cowansville, Pa. | 6/18 | E |
| 235 | US45SC 68141 | Marshaltown, Iowa | 6/13 | Boxholm, Iowa | 6/15 | W |
| 236 | US45SC 68347 | Indiana, Pa. | 6/10 | Gibsonia, Pa. | 6/19 | E |
| 237 | US45SC 68302 | Not released | No record | | | |
| 238 | US45SC 68388 | Ames, Iowa | 6/12 | Not reported | | |
| 239 | US45SC 68348 | Wadsworth, Ohio | 6/10 | Northfield, Ohio | 6/11 | E |
| 240 | US45SC 68349 | Altoona, Pa. | 6/10 | State College, Pa. | 6/10 | E |
| 241 | US45SC 68387 | Dunlap, Iowa | 6/12 | Dow City, Iowa | 6/14 | W |
| 242 | US45SC 68386 | Not reported. | | | | |

## EXPERIMENT IV

## The East-West Experiment, June 1945

As previously stated, the orientation of the magnetic-vertical lines with respect to the Coriolis lines produce regions where the gridwork character tapers off and a pattern of irregular parallelism or tangency occurs. Reference to Map II indicates this relationship over a fairly wide strip of land through the region which includes the states of Wisconsin, Illinois, Indiana, Kentucky and continuing in a south-south-east direction. An irregular line including all the points of tangency in this region will subsequently be referred to as "the line of tangency."

It seems evident that if birds are guided to a unique point (their home or conjugate point) by conditions requiring grid representations of the fields of influence, they would lack that guidance in a region where the indicating lines approximate a condition of parallelism. Such has been thought to be the case by racing pigeon men in the region just referred to.[16] The procedures of

---

[16] F. J. Sauerteig, "Indianapolis," Am. Racing Pigeon News (November 1942).

TABLE VII. Showing classifications of birds with regard to their releases east and west of line of tangency (including those released within 100 miles of this line).

| Yellow band No. | U. S. Army Signal Corps band No. | Place of release | Date | Place found | Date | +X | ±Y |
|---|---|---|---|---|---|---|---|
| | | | Group (1) Eastern releases | | | | |
| 211 | US45SC 68368 | Wadswörth, Ohio | 6/10 | State College, Pa. | 6/20 | +197 | + 0 |
| 215 | US45SC 68389 | Altoona, Pa. | 6/10 | State College, Pa. | 6/10 | + 33 | 0 |
| 216 | US45SC 68354 | Tiffin, Ohio | 6/11 | Hartville, Ohio | 6/14 | +106 | − 7 |
| 223 | US45SC 68372 | W. of Youngstown, O. | 6/10 | Loyalhanna, Pa | 6/11 | + 73 | −40 |
| 227 | US45SC 68362 | Butler, Pa. | 6/10 | State College, Pa. | 6/10 | +107 | 0 |
| 228 | US45SC 68381 | Youngstown, Ohio | 6/10 | State College, Pa. | 6/11 | +146 | 0 |
| 229 | US45SC 68385 | Indiana, Pa. | 6/10 | Johnstown, Pa | 6/12 | + 10 | −20 |
| 230 | US45SC 68363 | Ottowa, Ohio | 6/11 | Weirton, W. Va. | 6/20 | +172 | −33 |
| 234 | US45SC 68393 | Butler, Pa. | 6/10 | Cowansville, Pa. | 6/18 | + 20 | − 2 |
| 236 | US45SC 68347 | Indiana, Pa. | 6/10 | Gibsonia, Pa. | 6/19 | − 51 | +61 |
| 239 | US45SC 68348 | Wadsworth, Ohio | 6/10 | Northfield, Ohio | 6/17 | + 11 | +20 |
| 240 | US45SC 68349 | Altoona, Pa. | 6/10 | State College, Pa. | 6/10 | + 33 | 0 |
| | | | | | TOTAL | $+X=908$ | $+Y=81$ |
| | | | | | | $-X=51$ | $-Y=102$ |
| | | | Group (2) Western releases | | | | |
| 201 | US45SC 68367 | Marshaltown, Iowa | 6/12 | 11 mi. East of Sioux Falls, S. D. | 6/14 | +165 | −140 |
| 203 | US45SC 68390 | Arlington, Neb. | 6/14 | Nebraska City, Neb. | 6/14 | − 8 | +57 |
| 204 | US45SC 68300 | Ames, Iowa | 6/12 | Tennant, Iowa | 6/25 | + 92 | + 4 |
| 206 | US45SC 68343 | Cedar Rapids, Iowa | 6/12 | Casey, Iowa | 6/14 | +143 | + 8 |
| 212 | US45SC 68369 | Carroll, Iowa | 6/12 | Stuart, Iowa | 6/24 | − 14 | +42 |
| 214 | US45SC 68380 | Dunlap, Iowa | 6/12 | Fontanelle, Neb. | 6/14 | + 45 | + 4 |
| 224 | US45SC 68353 | Carroll, Iowa | 6/12 | Coon Rapids, Iowa | 6/14 | − 4 | +12 |
| 226 | US45SC 68391 | Waterman, Ill | 6/11 | Mt. Ayr, Iowa | 6/14 | +280 | +45 |
| 231 | US45SC 68377 | Waterman, Ill. | 6/11 | Chariton, Iowa | 6/12 | +232 | +30 |
| 232 | US45SC 68273 | Arlington, Neb | 6/12 | Garland, Neb. | 6/26 | + 41 | +22 |
| 235 | US45SC 68141 | Marshaltown, Iowa | 6/12 | Boxholm, Iowa | 6/15 | + 52 | −20 |
| 241 | US45SC 68387 | Dunlap, Iowa | 6/12 | Dow City, Iowa | 6/14 | 7 miles | |
| | | | | | TOTAL | $+X=1050$ | $+Y=224$ |
| | | | | | | $-X=26$ | $-Y=160$ |
| | | | Group (3) Birds released within 100 miles of line of tangency | | | | |
| 202 | US45SC 68272 | Joliet, Ill. | 6/11 | LaSalle, Ill | 6/16 | + 50 | +12 |
| 205 | US45SC 68345 | Morrison, Ill. | 6/11 | Marion, Ind. | 6/14 | +220 | −58 |
| 209 | US45SC 68355 | Morrison, Ill. | 6/11 | Gridlev, Ill. | 7/3 | + 47 | −58 |
| 233 | US45SC 68394 | Valparizo, Ill. | 6/11 | Kankakee, Ill. | 6/13 | + 30 | +18 |
| | | | | | TOTAL | $+X=347$ | $+Y=+30$ |
| | | | | | | $-X=-0$ | $-Y=-116$ |

this experiment were designed to test this phase of the theory.

On April 23, 1945, fifty five-week old youngsters were received from Fort Sam Houston, Texas. Pilotage and navigation training of these birds was started immediately. Their loft was not mobiled and training releases were made twice daily. These consisted of one, two, four, eight, sixteen, and twenty-five mile flights for pilotage experience, and many thirty-, fifty-, and seventy-mile flights, in several directions, for navigational training. On some of the longer flights they were liberated in twos, threes, and fives to make them more self-reliant and capable of independent navigational flights.

On June 10, 1945, when the experiment was to begin, there remained forty-two birds. The other eight were lost in the longer training flights, especially when released in small groups. They were transported by truck from State College to the points of liberation which were about equally spaced between State College, Pennsylvania, and Kearney, Nebraska. Table VI shows the time and place these birds were released as well as the time and place each was found and reported. Table VII shows the release and terminal points of all the birds whose flight terminals were determined. These include all reported by telegraph and in the case of the eastern releases, five which actually arrived at the home loft. In addition it classifies them as follows:

- Group 1 Those birds released east of the line of tangency (or region of confusion).
- Group 2—Those birds released west of the line of tangency.
- Group 3—Those birds which were released within a distance of one hundred miles east or west of the line of tangency and flew across it.

The column at the right on Table VII also shows the distances which each bird flew on a direct line toward the conjugate for the "western" releases and toward the home loft in the case of the "eastern" releases. The $Y$ column shows the perpendicular distances flown either (+) counter-clockwise or (−) clockwise from the direct line to conjugate or home point, respectively.

Table VIII shows the computation of results from the data on Table VII. The algebraic sums of the $X$ and $Y$ components of both groups were divided by the number of flights contributing to their totals. These values of ($\Sigma X$/No. flights) and ($\Sigma Y$/No. flights) were then added vectorily to determine the resultant "flight vector" for the group of birds involved.

As shown, the Eastern group (No. 1) on the average flew 1.8 miles in a counter-clockwise direction while averaging 71.4 miles toward the home (a deviation of approximately 1.0°) whereas the Western group (No. 2) deviated 5.8 miles in a clockwise direction while averaging 93.1 miles toward the conjugate point (a deviation of approximately 3.6°).

At the conjugate end of the experiment the birds, even without errors caused by large local magnetic anomalies, would be subject to errors of fifteen or twenty-five miles from the point, because of the natural errors of navigation to which they are subject. At the home end the pilotage training permits the bird, by means of local landmarks, to finish the last twenty-five or so miles of the journey to the loft. This probably accounts for the smaller angle of deviation for the eastern group.

Birds of group (No. 3) were deliberately liberated within the central two hundred-mile band, which should be a region of confusion insofar as the theory is concerned. It seems significant that the only birds to cross over the line of tangency were released in this band. More data will be obtained to check this indication.

## EXPERIMENT V

### Second Kearney Experiment, June 1945

This experiment deals with approximately two hundred homing pigeons trained at State College, Pennsylvania. In general, the preparations and procedures were the same as described in the previous experiment at Kearney, Nebraska. However, since the results of the previous tests had so thoroughly dovetailed with the requirements of magnetic-vertical Coriolis theory, at the suggestion of Major Otto Meyer and Lt. Arthur M. Lehman of the Army Signal Corps, it was decided to put a handicap in the way of these birds in order to give the theory the acid test. The plan involved training the birds by releasing them always north-west of State College, but when testing them for the conjugate point in the Kearney, Nebraska, region, making the releases well toward the north-east. In many of the pigeon races greater speed results from having the birds fly and refly over familiar territory and many believe the birds' ability to return depends on visual cues resulting from this careful training in the same direction.[17] Therefore, the significance of the above plane is self-evident.

During the first Kearney experiment in July 1944, five-foot meteorological balloons were used at State College and in Nebraska in an attempt to aid the birds in piloting the last ten or fifteen miles to the loft. The strong Nebraska winds, however, quickly blew them down and broke them so that their value was not determined. In this experiment eight-foot orange-colored box-kites were used for a similar purpose. The kites were flown during most of the training flights to the home loft at State College and during the two days the birds were released in Nebraska. Since no birds actually found the lofts this year, it is assumed that their effect was negligible.

---

[17] Ralph Gundlach, "Field study of homing pigeons;" Lucian H. Warmer, "The present status of the problems of orientation and homing by birds;" "Homing pigeons a puzzle to science," Am. Weekly (June 3, 1945).

TABLE VIII. Computations with diagrams of results.

Group (1)
Eastern releases

Distance traveled toward home loft $X_e = +908$ miles (11 birds)
Distance traveled away from home loft $-X_e = -\ 51$ miles (1 bird)
$\Sigma X_e\ \ +857$ miles (12 birds)
Deviation to left from the $X$ direction $Y_e = +\ 81$ (2 birds)
$-Y_e = -102$ (5 birds)
$Y_e = \ \ 0$ (5 birds)
$\Sigma Y_e = -\ 21$ (12 birds)

$$\frac{\Sigma X_e}{\text{No. birds}} = \frac{+857}{12} = 71\ 4 \text{ miles}; \quad \frac{\Sigma Y_e}{\text{No. birds}} = \frac{-21}{12} = -1.8 \text{ miles}$$

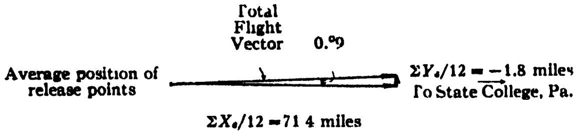


FIG. 2.

Group (2)
Western releases

Distance traveled toward conjugate point $+X_w = +1050$ miles (8 birds)
Distance traveled away from conjugate point $-X_w = -\ \ 26$ miles (3 birds)
$\Sigma X_w = +1024$ miles (11 birds)
Deviation to left from $X_w$ direction $Y_w = +\ 224$ miles (9 birds)
Deviation to right from $X_w$ direction $-Y_w = -\ 160$ miles (2 birds)
Net (to left) deviation from $X_w$ direction $\Sigma Y_w = +\ \ 64$ miles (11 birds)

$$\frac{\Sigma X_w}{\text{No birds}} = \frac{+1024}{11} = 93.1 \text{ miles}; \quad \frac{\Sigma Y_w}{\text{No. birds}} = \frac{+64}{11} = +5.8 \text{ miles}$$

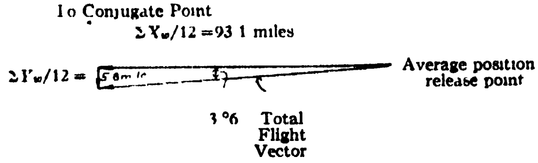


FIG. 3.

### Procurement and Training of Birds

Of the two hundred homing pigeons trained for this experiment one hundred fifty-nine were six-week old youngsters supplied by the Army Service Forces from the Army breeding lofts at Fort Sam Houston, Texas. The remainder were birds secured from various racing pigeon men a year previous and held over to be used as two-year-olds.

All the birds were housed in two mobile lofts having tall peaked roofs, six-foot ceilings, and six- by twelve-foot floor spaces. Each contained one hundred and twenty nest boxes with adequate avaries for sunning and bathing. The lofts were mounted on standard farm wagons having Timken high speed bearings, 6:00×16 automobile tires, and were painted a bright yellow. They were moved one hundred yards twice daily for receiving the birds from their training flights. The latter change in loft position was accomplished with a Fordson farm tractor.

Training flights were carried out morning and afternoon. After the initial short flights to the east, north, and west, to take care of their error in navigation, the birds were trained only from the north-west. Except for a few stormy days, training was continuous from the time the younger birds were received on April 14 until the day of departure for Kearney on June 10.

The navigation flights were extended to seventy miles airline. The birds flew the latter distance many times and on several occasions, when the wind velocity was practically zero, they homed in approximately seventy minutes.

## Table IX.

$X$ = distance bird flew toward conjugate point; $-X$ = away from point.
$Y$ = distance bird flew to left; $-Y$ to right of line drawn from liberation point to conjugate point. Released by: E-L = English-Lehman, H-B = Hendrickson-Brewer, H-R = Hendrickson-Rendle.

| Yellow celluloid band No. | U. S. Army or other seamless band No. | | Color | Release date | Release by | Place released | Date found | Place found | ±$X$ | ±$Y$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Group A | | | | |
| 6 | US45SC | 68240 | D.C. | 6/17 | E-L | Silver Creek, Neb. | 6/20 | Hampton, Neb. | + 22 | + 27 |
| 8 | US45SC | 68247 | DCWF | 6/16 | E-L | Silver Creek, Neb. | 6/18 | Spalding, Neb. | + 22 | − 40 |
| 22 | AU44 | W31805 | G.riz | 6/17 | H-R | Hampton, Neb. | 6/18 | Southerland, Neb | +166 | − 37 |
| 26 | US45SC | 68241 | R.C. | 6/17 | E-L | Silver Creek, Neb. | 6/19 | Chapman, Neb. | + 32 | + 8 |
| 33 | US45SC | 68285 | D.C. | 6/16 | E-L | St. Edward, Neb. | 6/18 | Trenton, Neb. | +187 | − 34 |
| 35 | US45SC | 68278 | B.CWF | 6/17 | E-L | St. Edward, Neb. | 6/18 | Stamford, Neb. | +132 | + 18 |
| 41 | US45SC | 68269 | D.C. | 6/16 | E-L | Silver Creek, Neb. | 6/18 | Bucklin, Kan. | +208 | +190 |
| 44 | AU44 | W31806 | N.C.O. | 6/17 | E-L | Albion, Neb. | 6/19 | Sargent, Neb. | + 51 | − 67 |
| 49 | AU44 | W31901 | N.C.O. | 6/17 | E-L | Albion, Neb. | 6/19 | Royal, Neb. | − 30 | − 37 |
| 53 | US45SC | 68316 | D.C. | 6/16 | E-L | St. Edward, Neb. | 6/19 | Stuart, Neb. | + 3 | − 98 |
| 54 | US45SC | 68357 | B.C. | 6/17 | E-L | Silver Creek, Neb. | 6/20 | Trenton, Neb. | +193 | − 7 |
| 56 | US45SC | 68249 | B.B. | 6/17 | E-L | Albion, Neb. | 6/20 | Eads, Col. | +333 | − 43 |
| 59 | US45SC | 68259 | B.C. | 6/16 | E-L | Albion, Neb. | 6/23 | Benkelman, Neb. | +200 | − 62 |
| 63 | US45SC | 68328 | Griz. | 6/16 | E-L | St. Edward, Neb. | 6/18 | Leigh, Neb. | − 32 | + 12 |
| 65 | US45SC | 68280 | B.C. | 6/17 | H-R | York, Neb. | 6/20 | Kearney, Neb. | + 78 | + 7 |
| 80 | US45SC | 68200 | B.C. | 6/16 | E-L | Albion, Neb. | 6/19 | Holbrook, Neb. | +142 | − 12 |
| 89 | US45SC | 68242 | CWF | 6/16 | E-L | McCook, Neb. | 6/18 | Des Moines, Iowa | +387 | − 75 |
| 91 | US45SC | 68312 | R.C. | 6/17 | E-L | Silver Creek, Neb. | 7/23 | Aurora, Neb. | + 29 | + 20 |
| 92 | US45SC | 68251 | R.C. | 6/16 | E-L | Silver Creek, Neb. | 6/17 | Wilson, Kan. | +112 | +137 |
| 94 | US45SC | 68155 | D.C. | 6/17 | H-R | York, Neb. | 6/19 | Concordia, Kan. | + 8 | + 90 |
| 115 | AU44V | 35135 | D.C. | 6/17 | H-R | Hampton, Neb. | 6/18 | Ellis, Kan. | +105 | +127 |
| 116 | AU44P | 48145 | R.C. | 6/17 | E-L | Silver Creek, Neb. | 6/26 | Ravenna, Neb. | + 68 | − 10 |
| 119 | AU43P | 11890 | BCWE | 6/17 | E-L | Silver Creek, Neb. | 6/19 | Hildreth, Neb. | + 98 | + 28 |
| 121 | AU43P | 11884 | B.C. | 6/17 | E-L | Silver Creek, Neb. | 7/19 | Salem, S. Dak. | − 76 | −142 |
| 122 | AU40P | 1799 | B.S. | 6/17 | E-L | Silver Creek, Neb. | 6/24 | Hebron, Neb. | + 29 | + 74 |
| 124 | US45SC | 68298 | B.C. | 6/17 | H-R | York, Neb. | 6/19 | Grand Island, Neb. | + 40 | − 7 |
| 130 | US44SC | 4163 | B.C. | 6/17 | E-L | Clarks, Neb. | 6/22 | Filley, Neb. | − 38 | + 87 |
| 134 | AU44V | 35136 | D.C. | 6/17 | E-L | Silver Creek, Neb. | 6/19 | Grand Island, Neb. | + 43 | + 9 |
| 136 | | — | D.C. | 6/16 | E-L | Silver Creek, Neb. | 6/20 | Wilsonville, Neb. | +152 | + 22 |
| 155 | AU43 | 11899 | B.C. | 6/17 | H-R | Hampton, Neb. | 6/18 | Steel City, Neb. | − 40 | + 63 |
| 157 | AU43 | Y157 | B.B. | 6/17 | H-R | Strombsurg, Neb. | 6/19 | Milford, Neb. | − 20 | + 30 |
| 160 | AU43 | Y160 | B.C. | 6/17 | E-L | St. Edward, Neb. | 6/19 | Linwood, Neb. | − 32 | + 39 |
| 162 | AU44T | 28634 | R.C. | 6/17 | H-R | York, Neb. | 6/18 | Leoti, Kan. | +199 | +145 |
| 167 | AU44V | 31005 | B.B. | 6/17 | E-L | Clarks, Neb. | 6/18 | Grand Island, Neb. | + 32 | + 9 |
| 170 | AU440 | 3139 | BCWE | 6/17 | H-R | Hampton, Neb. | 6/19 | Deshler, Neb. | − 4 | + 55 |
| 171 | AU45W | 31905 | B.B. | 6/17 | C-L | St. Edward, Neb. | 6/19 | Wahoo, Neb. | − 36 | + 60 |
| 183 | AU44V | 31012 | B.B. | 6/17 | E-L | Silver Creek, Neb. | 6/18 | Pierce, Neb. | − 32 | − 53 |
| 188 | US44 | 21158 | D.D. | 6/17 | H-R | York, Neb. | 6/20 | York, Neb. (22 m. N.W.) | + 18 | − 12 |
| 190 | US45SC | 68258 | B.C. | 6/17 | H-R | York, Neb. | 6/20 | Mullen, Neb. | +173 | − 93 |
| 192 | AU44V | 31017 | S.B. | 6/17 | H-R | York, Neb. | 6/25 | Cody, Neb. | +180 | −155 |
| 193 | AU43X | 27255 | B.C. | 6/17 | H-R | York, Neb. | 6/23 | Claflin, Kan. | + 60 | +160 |
| 197 | US45SC | 68248 | B.B. | 6/17 | E-L | Silver Creek, Neb. | 7/19 | Ft. Logan, Colo. | +395 | − 73 |
| | | | | | | | | | $\Sigma X = +3557$ | $\Sigma Y = +360$ |
| | | | | | | Group B | | | | |
| 47 | US45SC | 68274 | BCWE | 6/16 | E-L | Silver Creek, Neb. | 6/28 | Topeka, Kan. | − 59 | +210 |
| 101 | AU44 | 24126 | B.B. | 6/17 | E-L | Clarks, Neb. | 6/28 | Delmont, S. Dak. | − 40 | −132 |
| 107 | AU44D | 32149 | D.C. | 6/17 | E-L | Clarks, Neb. | 7/3 | North Riverside, Ill. | −440 | +140 |
| 135 | AU44V | 35174 | B.B. | 6/17 | E-L | Silver Creek, Neb. | 6/27 | Shelby, Neb. | − 7 | + 12 |
| 151 | US44SC | 20523 | B.C. | 6/17 | E-L | Albion, Neb. | 7/30 | Bridgewater, Iowa | −100 | +150 |
| 152 | US45SC | 68205 | B.B. | 6/17 | H-R | Aurora, Neb. | 6/29 | Portis, Kan. | + 43 | + 90 |
| 158 | AU45W | 31904 | B.B. | 6/17 | H-R | Stromsburg, Neb. | 7/14 | Ashton, Iowa | −115 | −100 |
| 163 | AU45W | Y163 | D.C. | 6/17 | E-L | Clarks, Neb. | 6/26 | Tilden, Neb. | − 16 | − 57 |
| | | | | | | | | | $\Sigma X = -734$ | $\Sigma Y = +313$ |

Training releases were made in groups of threes and fives a few times, but in most cases the birds were flown in groups of fifty.

On the morning of June 9th the birds were given a seven-mile exercise flight from the northwest. Immediately following, one of the lofts was transferred from its farm wagon base to a two-ton International Truck owned by the Agri-

TABLE IX.—*Continued.*

$X$ = distance bird flew toward conjugate point; $-X$ = away from point.
$Y$ = distance bird flew to left; $-Y$ to right of line drawn from liberation point to conjugate point. Released by: E-L = English-Lehman, H-B = Hendrickson-Brewer, H-R = Hendrickson-Rendle.

| Yellow celluloid band No. | U. S. Army or other seamless band No. | | Color | Release date | Release by | Place released | Date found | Place found | ±X | ±Y |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Group C | | | | |
| 10 | AU44 | 1475 | D.C. | 6/16 | H–B | McCook, Neb. | 6/16 | McCook, Neb. | 0 mile | |
| 19 | US45SC | 68287 | R. | 6/17 | H–R | Stromsburg, Neb. | 6/24 | Stromsburg, Neb. | 0 mile | |
| 32 | US45SC | 68289 | S.B. | 6/16 | E–L | Silver Creek, Neb. | 6/22 | Genoa, Neb. | 10 miles | |
| 73 | US45SC | 68295 | D.D. | 6/19 | H–R | Stromsburg, Neb. | 6/22 | Stromsburg, Neb. | 0 mile | |
| 78 | US45SC | 68297 | B.C. | 6/16 | H–B | McCook, Neb. | | McCook, Neb. | 0 mile | |
| 132 | US45SC | 68157 | B.C. | 6/17 | E–L | Silver Creek, Neb. | 6/22 | Genoa, Neb. | 7 miles | |
| 141 | US44SC | 68275 | R.C. | 6/17 | E–L | St. Edward, Neb. | 9/19 | Genoa, Neb. | 10 miles | |
| 153 | AU44V | 31015 | B.C. | 6/17 | H–R | Hampton, Neb. | 6/21 | Aurora, Neb. | 7 miles | |
| 199 | AU44T | 28631 | B.C. | 6/17 | H–R | Stromsburg, Neb. | 6/21 | Osceola, Neb. | 10 miles | |

cultural Engineering Department of the College. The second was coupled to the back of the truck. All feed, kites, and other needed equipment was put under a false floor built in the latter loft.

On June 10th at 7 A.M. the trip to Kearney, Nebraska, was started. The party consisted of Roy Johnson, in charge of trucks, Jack Nesbitt, in charge of kites, Donald Brewer, in charge of birds, Major Otto Meyer, Lt. A. M. Lehman, P. F. English, and H. L. Yeagley. The trip of more than twelve hundred miles was made in three days. Two men, including the driver, handled the truck, and drivers were changed every two hundred or so miles, while the remainder of the party rode in a car belonging to the Pennsylvania Cooperative Wildlife Research Unit of the Pennsylvania State College and driven by Dr. English. At Ames, Iowa, Dr. George Hendrickson joined the party with a car owned by the Iowa Cooperative Research Unit. Mr. J. A. Rendle of Cozad, Nebraska, also joined the party at Kearney and assisted in the experimental procedures.

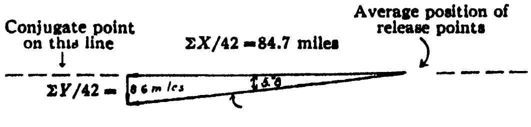
Upon arrival at Kearney the lofts were placed one and twenty miles, respectively, north of the town. This spaced them about equally north and south of the western State College conjugate. After one day for resting the birds, they were taken in groups and released on an average of seventy-five miles to the N.N.E. and E.N.E., and N.E. of the loft locations.

On June 15th at ten P.M. reports from the United States Weather Bureau indicated that the next day would be clear with scattered clouds and that wind velocities would average ten miles an hour or lower. About four o'clock the next morning (June 16th) Dr. English and Lt. Lehman took one hundred eighteen birds for liberation at points averaging seventy-five miles to the north-east of the conjugate. The birds were all released during the morning groups of threes and fives in the regions around the

TABLE X.

Group A
(Including al birds reported within 10 days of liberation)

$\Sigma X$ for 42 birds = +3557 miles
$\Sigma X/42 = 84.7$ miles
$\Sigma Y$ for 42 birds = +360 miles
$\Sigma Y/42 = 8.6$ miles

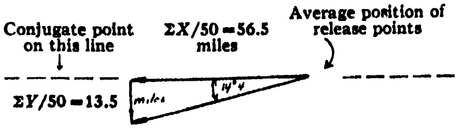

Conjugate point on this line
$\Sigma X/42 = 84.7$ miles
Average position of release points
$\Sigma Y/42 =$
8 6 miles
5.8

Total Flight Vector

FIG. 4.

Groups A and B
(Including all birds reported up to December 24, 1945)

$\Sigma X$ for 50 birds reported = 2823 miles
$\Sigma X/50 = 56.5$ miles
$\Sigma Y$ for 50 birds reported = 673 miles
$\Sigma Y/50 = 13.5$ miles


Conjugate point on this line
$\Sigma X/50 = 56.5$ miles
Average position of release points
$\Sigma Y/50 = 13.5$
miles
14°4

Total Flight Vector

FIG. 5.

TABLE XI. Showing flight vectors, $X$ and $Y$ distances flown $\Sigma X$ and $\Sigma Y$ values and "total flight vector" for all Army pigeons reported within ten days of liberation. Number reported—20.

$\Sigma X = +2664$ miles $\Sigma Y = -45$ miles

$\Sigma X/20 = +133.2$ miles $\Sigma Y/20 = -2.3$ miles

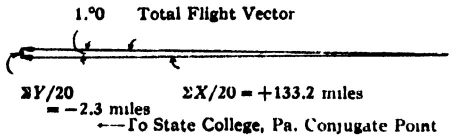


FIG. 6.

| Band number | $X$ | $Y$ |
|---|---|---|
| 6 | + 22 | + 27 |
| 8 | + 22 | + 40 |
| 26 | + 32 | + 8 |
| 33 | +187 | − 34 |
| 35 | +132 | + 18 |
| 41 | +208 | +190 |
| 53 | + 3 | − 98 |
| 54 | +193 | − 7 |
| 56 | +333 | − 43 |
| 59 | +200 | − 62 |
| 63 | − 32 | + 12 |
| 65 | + 78 | + 7 |
| 80 | +142 | − 12 |
| 89 | +387 | − 75 |
| 91 | + 29 | + 20 |
| 92 | +112 | +137 |
| 94 | + 8 | + 90 |
| 124 | + 40 | − 7 |
| 190 | +173 | − 93 |
| 197 | +395 | − 73 |
| | $\Sigma X = +2664$ mi. | $\Sigma Y = -45$ miles |

Nebraska towns of Clarks, Silver Creek, St. Edward, and Albion. Simultaneously Dr. Hendrickson and Donald Brewer released ten birds to the south-west (near McCook, Nebraska) for control purposes. As indicated by the weather reports, the day was unusually good for bird navigation.

The next morning, June 17th, with weather report and actual weather similar to that of the preceding day, Dr. Hendrickson and Mr. Rendle took the fifty-six remaining birds for liberation to regions around Stromsburg, York, and Hampton, Nebraska. The distance was about the same as the preceding day but the direction was more toward the east. Again the birds were released well before noon in groups of threes and fives. As in previous experiments the birds carried messages in Army message carriers asking the finder to report when and where the bird was found.

### *Results*

Table IX lists all birds which were reported (thirty-two percent of total released) up to the present writing. It lists the time and place of release, persons in charge of liberation, time and place they were found, and the distances flown along and at right angles to a line drawn through the conjugate point and the release point.

From racing-pigeon men it has been learned that birds which do not arrive home from one- and two-hundred-mile flights within a few days are likely not to arrive at all. This is taken as an indication that they become discouraged or fly at random after that time. For this reason all birds reported more than ten days after liberation were listed separately (Group B, Table IX) and this data used in a subsequent calculation to test its effect on the more reliable early data. A second separate listing consists of birds which never left, or perhaps returned to the release areas (Group C, Table IX). In any case, they were reported found within a ten-mile radius of the release point. The complete listing of the birds on Table IX is as follows:

Group A Birds reported within a nine- or ten-day period after liberation.

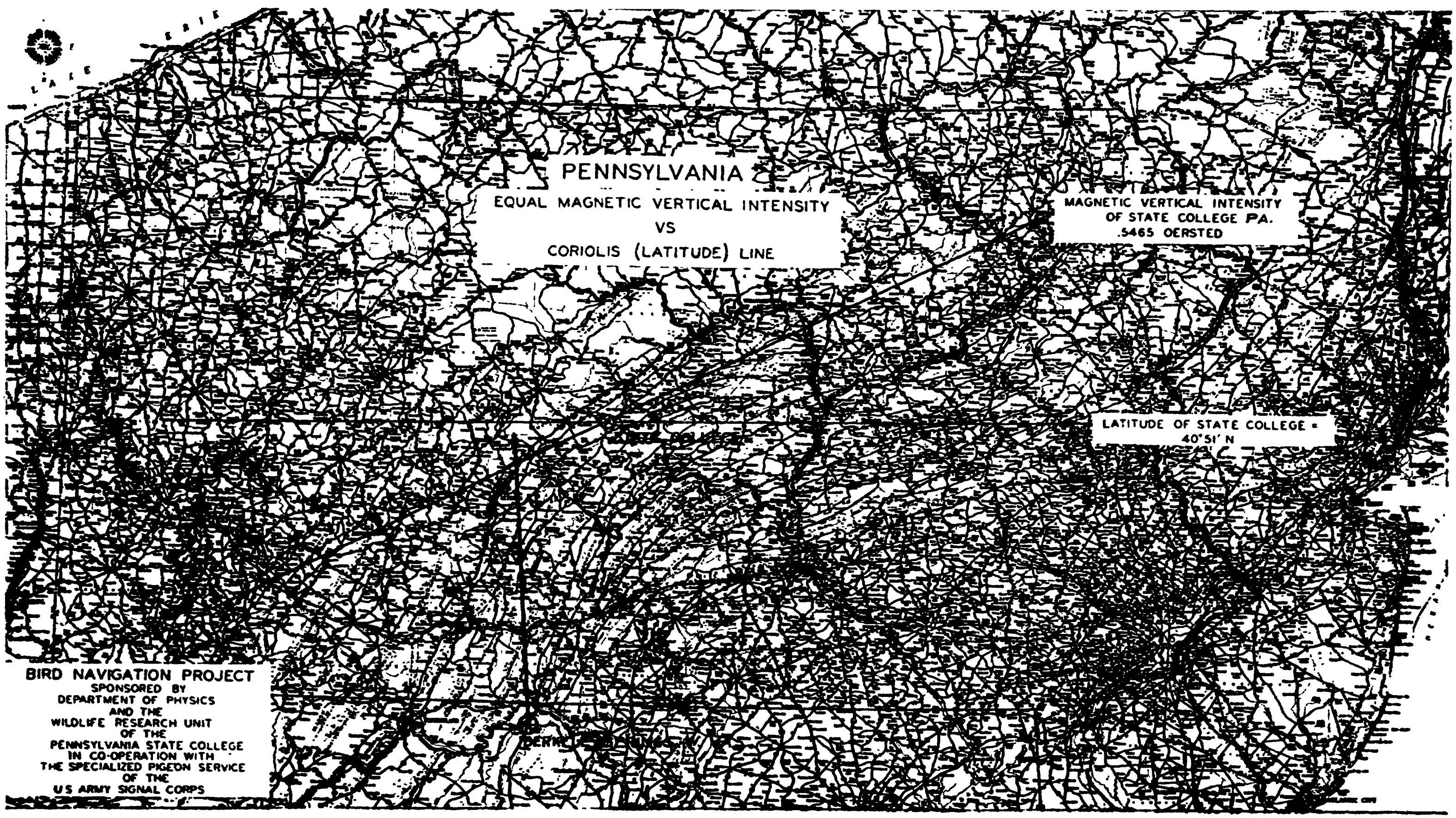
Group B Birds reported after June 26th.

Group C Birds which were reported less than ten miles from the release point.

In working out and interpreting the data of this experiment the same plan was followed as with the previous ones. The flight (release point to flight terminal) of each bird was plotted and a straight line drawn from the release point to the conjugate point. The algebraic sums $\Sigma X$ and $\Sigma Y$ of the components of all the flights toward and at right angles to the theoretically required direction line were divided by the number of birds and added vectorily to obtain the "total flight vector" for the groups.

On Table X is shown the average angle of deviation obtained from the flight data of the birds listed in Table IX. Those of Group A show a counter-clockwise deviation of 5.8°. When the data from those reported later (Group B) are used in the computations, the average deviation is seen to be +13.5 degrees.

As in the case of the 1944 Kearney experiments, the experimental birds seemed to behave



Map IV. Map of Pennsylvania with northern parts of Virginia and West Virginia. Arrow represents total flight vector for reported birds of Special release No. 3 on Table III.

according to the magnetic vertical-Coriolis theory by flying on the average, with only small deviation from a line toward the computed conjugate point. It is noteworthy that the angle of deviation for the birds reported within reasonable time limits is extremely small (5.8°), and this in spite of the fact the birds were trained in a manner devised to confuse them if possible. If the training flights around State College, Pennsylvania, had misled them into flying toward the sun or the south-eastern sky each day, then in the Nebraska liberations they would have flown toward the state of Missouri. This would have been roughly ninety degrees from both State College, Pennsylvania, and Kearney, Nebraska, directions.

During the study and computation of the data in this report it was noticed that the pigeons supplied by the Army Signal Corps, on the average, made more accurate flights in the direction of the conjugate point than those obtained from private sources. It therefore seemed desirable that a separate computation be made for these birds. The summation of their flight components ($\Sigma X, \Sigma Y$) and average components ($\Sigma X$/No. birds, $\Sigma Y$/No. birds) are:

Algebraic sum of distance flown in the direction of the conjugate by the twenty Army birds reported up to June 26, $\Sigma X = 2664$ miles.

Algebraic sum of distance flown at right angles to the above direction, $\Sigma Y = -45$ miles.

As indicated on Table XI, this represents an average flight of 133.2 miles in the $X$ direction and 2.2 miles in the $-Y$ direction or a deviation of only one degree from the average direct line to the conjugate point. This seems to indicate that the Army birds are superior navigators. The small angle of deviation also seems remarkable in view of the flatness of the Nebraska country as contrasted with the mountains in Pennsylvania where the birds were trained, and the anomalous set-up of the earth's magnetic field through which they had to fly in both regions.

## REMARKS

The fundamental importance of the foregoing explanation of the long mysterious bird navigation problem and the experimental data obtained is fourfold:

1. The results furnish an experimental indication that a bird moving across a magnetic field experiences an awareness to the effect.
2. The age-old mystery of bird migrations may be clarified by the results of this and already planned future experiments with wild birds. The ability of the Golden Plover[18] to navigate from Hawaii to Alaska, 3000 miles one way over ocean vastness, may possibly have an explanation in the homing pigeon's technique.
3. The biologists and psychologists can make use of its findings in studying the effect of a moving magnetic field on living nerves, an effect lacking demonstration up to now. Part of the ills occurring to pilots in super-speed planes and other high velocity craft, and attributed entirely to large values of acceleration, may conceivably be due to moving rapidly through the earth's magnetic field
4. There is evidence that wild birds and homing pigeons become confused when in the region of powerful broadcasting stations. Although there is much controversy on the subject, the evidence included in this report, both from the standpoint of the magnetic wing and the conjugate point experiments, indicates that the phenomenon could exist. Further evidence of this is also being sought with new experiments on homing pigeons now under way at The Pennsylvania State College.

## ACKNOWLEDGMENTS

The experimental work has been supported by: the Department of Physics and The Department of Zoology and Entomology of The Pennsylvania State College; the Pennsylvania Wildlife Research Group at The Pennsylvania State College and the Army Service Forces of the United States Army Signal Corps.

The men mainly responsible for carrying out the experimental procedures are: Dr. H. L. Yeagley, Department of Physics, and Professor Merrill Wood, Department of Zoology, both at The Pennsylvania State College; Dr. P. F. English and Dr. Logan Bennett, both of the Pennsylvania Wildlife Research Unit; Dr. George Hendrickson, Associate Professor of Wildlife Management, Iowa State College, and Major

---

[18] F. C. Lincoln, The Migration of American Birds, p. 54.

Otto Meyer, in charge of the Pigeon Service Agency, Army Service Forces, U. S. Army Signal Corps, during World War II.

Note: Illustrative maps and tables are the same as those used in the Reports to the Army Service Forces.

## BIBLIOGRAPHY

G. C. Aymar, *Bird Flight* (Garden City Publishing Company, Garden City, New York, 1935), p. 99.

C. E. Brown, "Homing pigeons exposed to radio frequency waves," Sci. Am. **160**, 45 (1939).

J. E. Brown, "The pigeon cooker," Radio News (April 1938).

J. Casamajor, "The mysterious 'seris de L'espace,'" Rev. Sci., Paris, **65**, 555 (1927).

E. Chamberlain, *The Homing Pigeon* (The Homing Pigeon Publishing Company, Manchester, England, 1907).

R. D. Chard, "Visual acuity in the pigeon," J. Exper. Psychol. **24**, 588 (1939).

H. E. Dodge, "The racing homer," Feathered World, London, (1922).

M. E. Duchatel, "Faculté de direction du pigeon voyageur," Ann. de Psychol. Zool. **1**, 22 (1901).

S. Exner, "Negative Versuchsergebnisse über das Orientierungsvermögen der Brieftauben," Sitz. Wien Akad. Wiss. **101**, 318 (1893).

Experiments with homing pigeons sensitivity to radio frequency waves, conducted by 285th Signal Pigeon Company, March 1945.

D. S. Farner, "The return of robins to their birthplaces," Bird Banding **16**(3), 81 (July 1945).

Dr. D. R. Griffin, "The Sensory Basis of Bird Navigation," The Quarterly Review of Biology, **19**, 1 (March 1944), pp. 15–31.

R. H. Gundlach, "A field study of homing in pigeons," J. Compar. Psychology. **13**, 397 (1932).

R. H. Gundlach, "A test of 'directional sense' in cats and pigeons," J. Compar. Psychol. **12**, 347 (1931).

J. Huie, "Orientation," Homing Pigeon **3**, 20 (1881). (Reprinted from J. Horticult.)

G. Ising, "The physical possibility of a biological sense of orientation based on the rotation of the earth," Ar. Math. Astrophys, 32A, No. 4 (1946).

Kingsley, "Orientation by inherited memory," Nature, **18** (June 1869).

Fredrick C. Lincoln, *The Migration of American Birds* (Doubleday Doran and Company, Inc., Garden City, New York, 1936).

W. M. Levi, *The Pigeon* (R. L. Bryan Co., Columbia, South Carolina. 1941).

O. H. Mowrer, "The nsytagmic response of the pigeon to constant angular acceleration at liminal and supraliminal intensities," Jour. Compar. Psychol. **19**, 177 (1935).

F. M. Packard, "The return of young robins to their birthplaces," Bird Banding **18**(1), 30 (January 1947).

J. L. Peters, *Check-list of Birds of the World* (Harvard University Press, Cambridge, Massachusetts, 1931), Vol. 1 or 2; *ibid.*, Vol. 2, 1937; *ibid.*, Vol. 3, 1940; *ibid.*, Vol. 4.

E. S. Starr, "Homing pigeons," Cent. Illust. Month. Mag. **31**, 361 (July 1886).

R. M. Strong, "On the olfactory organs and the sense of smell in birds," J. of Morphol. **22**, 619 (1911).

F. J. Sauerteig, "Indianapolis," American Racing Pigeon News, November 1942, p. 1.

G. Reynaud, "The laws of orientation among animals," Smith, Inst. Ann. Report of Board of Directors, p. 481 (translated from Rev. des Deux Mondes 146) 380 (1898), also O. H. Mowrer, J. Compar. Psychol. 19, 177 (1935).

W. H. Thorpe and D. H. Wilkinson, "Ising's theory of bird orientation," Nature, December 21 (1946).

A. Thauzies, "L'orientation," Rev. Sci. **9** (series 4), 392 (1898).

A. Thauzies, "L'orientation du pigeon voyageur: donnes experimentales," Rev. Sci. **2** (series 5), 417, 453 (1904).

C. Viguier, "Le sense de l'orientation," Rev. Phil. **14**, 1–36 (1882).

G. J. Van Oordt, *Bird Migration* (Vogeltrek) (E. J. Brill) Leiden, Netherlands, 1944); Review in Bird Banding, **18** (1), 48 (1947).

J. B. Watson, "Recent experiments with homing birds," Harper's Month. Mag., 457 (May 1913).

J. B. Watson, and K. S. Lashley, "An historical and experimental study of homing," J. Carnegie Inst. of Washington (1915).

A. G. Webster, *Dynamics* (G. E. Stechart and Company, New York), second edition, pp. 320–322.

K. Wodzicki, et al., "Experiments on homing in birds," Nature, **35** (Jan. 1, 1938).

# Inversion Applied to the Solution of 3-Dimensional Electromagnetic Problems*

A. Bloch
*Research Laboratories, The General Electric Company (Ltd.), Wembley, England*

(Received March 18, 1947)

This paper presents an extension of the method of inversion to certain electromagnetic problems and rectifies at the same time an error of long standing in the solution of an elementary problem which seems to have acted as a bar to just this extension.

If the image of a charge in a sphere is derived by the method of inversion, as is done in this paper, it becomes immediately obvious that the image of radially oriented dipole is another radial dipole plus a free charge proportional to the moment of the dipole. The potential caused by the free charge is not of negligible magnitude, and this indicates that the solution given by Sir J. Larmor and independently by S. P. Thompson and Miles Walker for the case of a magnetic dipole outside a sphere of infinite permittivity cannot be correct, for the image used therein consists of a simple dipole. However, it can be shown in addition that the latter image is not void of physical significance: the vector potential of such an image transforms in a simple manner and from this in turn follow simple laws of transformation for flux linked with loops and for mutual inductance between such loops.

This extension of the method of inversion is used in the second part of the paper to deal with a problem complementary to that of Larmor-Thompson, namely, with that of a spherical wall impermeable to magnetic flux. This problem is of practical importance as an idealization of an inductor used at radiofrequencies and placed inside a metal screen, and it has for this reason been dealt with by direct attack on several previous occasions. It is shown here that the action of such a screen on the field produced by the loop can be replaced by the introduction of a simple image loop. From this follows then a simple way of calculating the change in self and mutual inductance produced by such a screen, even in cases more general than considered heretofore. The losses caused by the screen can also be easily calculated by the use of H. A. Wheeler's concept of equivalent magnetic skin thickness.

## SYMBOLS

$O$ = center of inversion
$K$ = constant of inversion
$P$ = point at which a charge or current element is situated
$P'$ = inverse of $P$ $[OP \times OP' = K^2]$
$N$ = point at which the value of a potential is considered
$N'$ = point inverse to $N$
$l = PN$
$l' = P'N'$
$r, r'$ = a distance from $O$ and its inverse
$q, q'$ = an electric or magnetic charge and its inverse
$m, m'$ = a magnetic moment and its inverse
$I, I'$ = an electric current and its inverse
$V$ = an electric or magnetic potential
$\mathbf{A}$ = Vectorpotential
$\phi$ = magnetic flux
$L$ = self-inductance
$M, M_{12}$ = mutual inductance.

## INTRODUCTION

THE purpose of this paper is to present an extension of the method of inversion to the solution of certain electromagnetic problems and at the same time to rectify an error of long standing which seems to have acted as a bar to just this extension. In 1888 Sir J. Larmor considered[1] the magnetic field caused by a circular current loop in the presence of a cocentric soft iron sphere and concluded that for infinite permittivity of this sphere its action on the field outside its boundary could be replaced by that of an "image" of the original loop in the sphere. An identical solution was given independently by Silvanus Thompson and Miles Walker[2] and is still referred to in modern textbooks. It does not seem to have been noticed so far that an image of this kind does not solve the problem given.

If these images are introduced by the method of inversion—as will be done hereafter—such failure will be immediately evident. At the same time it is then not difficult to see that such an image obeys a simple law of transformation for its vector potential, from which follow in turn simple laws of transformation for magnetic flux through loops and mutual inductance between loops.

---

* Communication from the Staff of the Research Laboratories of The General Electric Company, Limited, Wembley, England.

[1] Sir J. Larmor, "Electromagnetic and other images in spheres and planes," Quart. J. Pure and App. Math., 1–8 24, 94 (1889).

[2] S. P. Thompson and M. Walker, "Mirrors of magnetism," Phil. Mag. **34** (5th series) 213–224 (1895).

This extension is dealt with in the first part of the paper, to be followed in the second part by an application to a problem complementary to that of Larmor and Thompson and Walker, namely, to that of a sphere impermeable to magnetic flux. This case arises when we study the reaction of a screening can at radio frequencies on an enclosed loop, as has been done on several occasions. H. Kaden[3] considered a loop of very small dimensions situated at the center of a spherical screen. C. F. Davidson and J. C. Simmonds[4] have extended this solution to the case of a loop of finite dimensions. H. Buchholz[5] has given solutions for the self- and mutual-inductance of loops which are parallel to each other and coaxially situated in any position along the radius of the screening sphere. All these authors use a direct method of attack and require, therefore, a mathematical apparatus of some complexity.

The present papers show that the action of the screen can be represented as the action of a simple image loop and thus reduces all the problems mentioned to the well solved determination of the mutual inductance between two parallel and coaxial loops. Actually, the methods of the paper are capable of dealing with some slightly more general arrangement of the loops and with certain non-circular loops.

## I. THE METHOD OF INVERSION AND ITS EXTENSION TO MAGNETIC IMAGES

Let $O$ be a fixed point (the *center of inversion*) and $P$ a variable one; determine on $OP$ a point $P'$ so that the distances $OP$ and $OP'$ fulfil

$$OP \times OP' = K^2 \tag{1}$$

($K$ = *constant of inversion*).

Then if $P$ describes a line, or a surface, or a volume, $P'$ will describe a corresponding geometrical element (the *inverse* or the *image* of the original element).

The most important properties of this well-known geometrical transformation are recalled by Figs. 1 and 2. Figure 1 shows that the inverse of a circle contained in a plane through the center of inversion is another circle in the same plane. Hence the inverse of a sphere must be another sphere, that of a circle in a general position—defined by the intersection of 2 spheres—must again be a circle, and the inverse of a plane (sphere through the infinitely distant point) a sphere through the origin. Figure 2 illustrates the "conformal" property of the transformation: the angles of intersection between lines and surfaces remain unchanged in the transformation.

To apply this method to the solution of problems of the theory of potential we have to add to this a law for the transformation of charges, as follows: A point charge of magnitude, say $q$, situated at a point, say $P$, shall be represented by another point charge of magnitude $q'$ situated at the image $P'$ of $P$, whereby

$$\frac{q'}{q} = \pm\frac{K}{OP} = \pm\frac{OP'}{K} = \left(\frac{OP'}{OP}\right)^{\frac{1}{2}}. \tag{2}$$

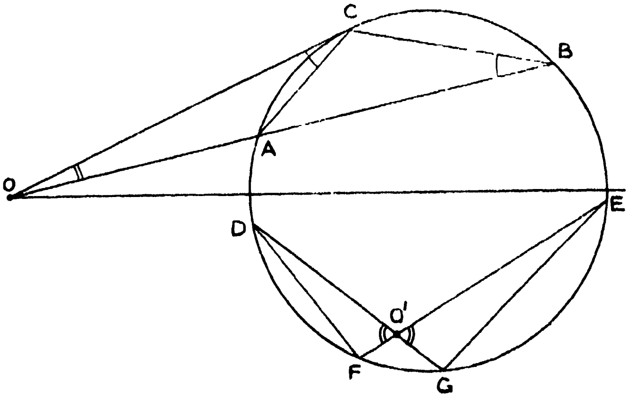


FIG. 1. The inverse of a circle is another circle. $\sphericalangle OCA = \sphericalangle CBA$ when $OC$ is tangent to circle; hence $\triangle OCA$ is similar to $\triangle OBC$, and $OA \times OB = OC^2$. If $K = OC$, this means that the circle inverts into itself; if $K$ is larger or smaller, the result of the inversion is correspondingly larger or smaller, but it will still be a circle. There is a similar proof if the center of inversion $O'$ is situated inside the circle: $\triangle FO'D$ is similar to $\triangle GO'E$ and $DO' \times O'G = FO' \times O'E =$ const.

---

[3] H. Kaden, E.N.T. **10**, 277 (1933).

[4] C. F. Davidson and J. C. Simmonds, W. Eng., **22**, 3 (1945).

[5] H. Buchholz, Archiv. f. Electrotechnik **28**, 556 (1934).

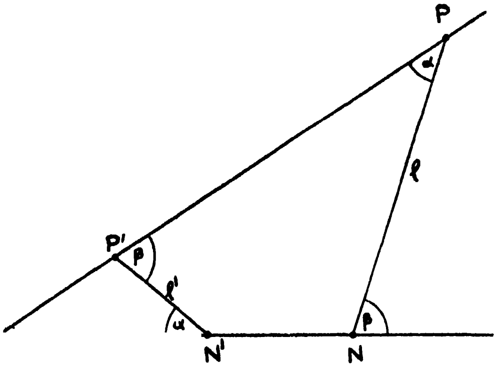


FIG. 2. If $OP\times OP'=ON\times ON'=K^2$, then $OP':ON'=ON:OP$, i.e., $\sphericalangle OP'N'=\sphericalangle ONP$. From this follows (a) if $l$ and $l'$ are infinitesimally small, the angles formed by these line elements with their radii vectorii are equal; (b) $l:l'=ON:OP'=OP:ON'=\left(\frac{ON\times OP}{ON'\times OP'}\right)^{\frac{1}{2}}=\frac{ON\times OP}{K^2}=\frac{K^2}{ON'\times OP'}$.

The reason for selecting this particular relation will become clear when we inquire into the potential $V$, caused by such a charge at some point $N$, and compare it with the potential $V'$ caused by the image charge at the image point $N'$ (see Fig. 2).

In this case $V=q/l$, and $V'=q'/l'$, hence

$$\frac{V'}{V}=\frac{q'}{l'}\cdot\frac{l}{q}=\pm\frac{K}{OP}\cdot\frac{OP}{ON'}$$

$$=\pm\frac{K}{ON'}=\pm\frac{ON}{K}=\pm\left(\frac{ON}{ON'}\right)^{\frac{1}{2}}. \quad (3)$$

The last equation shows that the transformed potential $V'$ can be calculated from the original potential without the need for inquiring into the position of the charge which "caused" the potential. Hence the potential will still transform accordingly to the same rule if it is due not to a single charge but to an assembly of charges.

These relations allow us to deduce the solution of a problem for the inverse space from the solution of the corresponding problem for the original space and *vice versa*. As an example we recall the case of a point charge $q$ in front of an infinite earthed plane. This problem is of course equivalent to that of the original charge, $q$, and of the mirror image, $q_m$, of this charge in this plane. By inversion from any point not coinciding with the plane the latter transforms into a sphere, with one charge $q'$ inside and the other one $q'_m$ outside its boundary. From Eq. (3) it is then seen that the potential of this sphere is again zero (for all points of the sphere not coinciding with the center of inversion the multiplier $K/ON'$ is finite; with a finite gradient of the field in the neighborhood of the center of inversion the potential of this point must then be zero too). Thus, the transformed arrangement solves the problem of a point charge inside or outside an earthed sphere. The "answers" for this case can be obtained either by appropriate translation from the results of the "plane" case, or, simpler still, by noting directly the fact that here too the total field is due to the combined action of the charge $q'$ and of the "image in the sphere" $q'_m$.

It is useful to note that we are led to this image charge also by a slightly different application of the method of inversion. If we invert the charge $q'$ and the sphere from the center of the sphere as inversion center in such a way that the sphere changes into itself (constant $K$ of inversion equal to the radius of the sphere) then $q'$ transforms into the image $q'_m$ previously found.*

From the fact that image and original charge are thus related by Eq. (2), it follows immediately, that in the case where we are given two charges of equal magnitude, forming a radially oriented dipole, the corresponding image charges cannot be equal. If the given charges are situated at, say $r$ and $r+dr$, the image charges are in this case situated at

$$r'=\frac{K^2}{r}, \text{ and } r'+dr'=\frac{K^2}{r+dr},$$

---

* That this is so—at least geometrically—is easily seen if we consider what happens to all those spheres which are common to the two charges $q$ and $q_m$ of the "plane" arrangement. These intersect normally with the plane, and they transformed therefor, in the first inversion into all those spheres which pass through the two charges $q'$ and $q'_m$ and intersect normally with the earthed sphere. In the second inversion—with the center of the sphere as inversion center—each of these spheres transforms into itself, i.e., their first common point must change into their second common point. (Their points of intersection with the earthed sphere remain fixed, and we know also that the angle of intersection must be again 90°.) Electrically the transformed arrangement could differ from the original arrangement then only by having both charges increased or decreased by the same factor. However, it is seen at once that this factor must be equal to unity as the electrical displacement at the surface of the sphere (charge per unit area) remained unchanged in this transformation (see Eq. 2)).

and they differ in magnitude by

$$\Delta q' = \frac{q \cdot K}{r^2} dr = \frac{m \cdot K}{r^2}, \tag{4}$$

where $m = qdr$ is the moment of the original dipole.

The case of an isolated sphere (with the given charges outside the sphere) differs from the case of an earthed sphere only insofar that a further charge at the center of the sphere is added so that the sum total of all the charges inside the sphere equals zero whilst the surface of the sphere still remains an equipotential surface.

The case of a magnetic dipole outside a shell of infinite permittivity is perfectly analogous to this latter case. From this follows immediately that a solution which takes the image of a radial dipole to be simply another dipole is not a case of ordinary inversion and is unlikely to follow the transformation law of Eq. (3) for the potential. (The potential produced at the surface of the sphere by the neglected free charge $\Delta q'$ is evidently of the same order as that produced by the dipole itself.)

This conclusion can be easily confirmed if we go to the trouble of writing down explicitly the expression for the magnetic potential of such a dipole. This expression, however, suggests further that it is now the magnetic vector potential of the dipole that follows the simple transformation law of Eq. (3). The magnetic vector potential of a radial dipole of moment $m$ situated at $P$ (Fig. 2) has at point $N$ the value

$$A = \frac{m}{l^2} \sin\alpha. \tag{5}$$

The direction of this vector is normal to the plane of the paper. The vector potential at $N'$ caused by the image dipole situated at $P'$ will be of magnitude

$$A' = \frac{m'}{l'^2} \sin\beta. \tag{6}$$

If we assume

$$\frac{m'}{m} = \frac{q' \cdot dr'}{q \cdot dr} = \frac{K}{OP} \cdot \frac{OP'}{OP} = \frac{K^3}{OP^3}, \tag{7}$$

we have

$$\frac{A'}{A} = \frac{m'}{m} \cdot \frac{l^2}{l'^2} \cdot \frac{\sin\beta}{\sin\alpha} = \frac{K^3}{OP^3} \cdot \frac{OP^2}{ON'^2} \cdot \frac{ON'}{OP'}$$

$$= \frac{K}{ON'} = \frac{ON}{K} = \left(\frac{ON}{ON'}\right)^{\frac{1}{2}}. \tag{8}$$

The foregoing finds an immediate application if we consider the fields produced by current loops. The magnetic effects of such a loop (not necessarily of circular shape), in which a current $I$ flows, is equivalent to that of a magnetic shell bounded by the loop, the strength of this shell (i.e., the magnetic moment per unit area) being equal to $I$.

Transformation of this shell produces as image another shell; if the original shell was spherical with center $O$, the image shell will also be spherical, and again of constant strength. Since corresponding elements of these two shells are related as to their surface areas as

$$\frac{S'}{S} = \left(\frac{r'}{r}\right)^2, \tag{9}$$

we have

$$\frac{I'}{I} = \frac{(m'/S')}{(m/S)} = \frac{K^3}{r^3} \cdot \frac{r^2}{r'^2} = \frac{K}{r'} = \frac{r}{K} = \left(\frac{r}{r'}\right)^{\frac{1}{2}}. \tag{10}$$

This means that we may replace this shell in turn by another current loop, the image of the original loop, the strength of the image current being given by Eq. (10).

The last equation may be used to show independently that the vector potential of such a loop transforms as stated in Eq. (8).

We have for the vector potential of the original loop

$$\mathbf{A} = \oint d\mathbf{A} = \oint \frac{I d\mathbf{s}}{l}, \tag{11}$$

and for that of the transformed loop

$$\mathbf{A}' = \oint dA' = \oint \frac{I' \cdot d\mathbf{s}'}{l'}. \tag{12}$$

The two contours over which these two line integrals are to be taken correspond to each other element $d\mathbf{s}$ for element $d\mathbf{s}'$, and the ratio

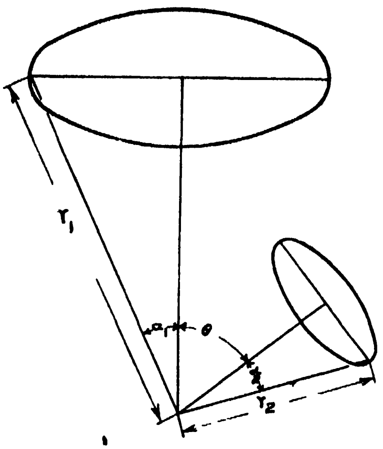


FIG. 3. The notation above $\gamma_2$ should read $\alpha_2$. The point at which the radii of the loops intersect should be labelled *O*.

of corresponding contributions $dA$ and $dA'$ is

$$\frac{d\mathbf{A}'}{d\mathbf{A}}=\frac{I'}{I}\cdot\frac{d\mathbf{s}'}{d\mathbf{s}}\cdot\frac{l}{l'}=\frac{K}{OP'}\cdot\frac{OP'}{OP}\cdot\frac{ON}{OP'}=\frac{ON}{K}. \qquad (13)$$

Hence the ratio of the two integrals themselves is given by the same constant.

From the transformation formula for the vector potential follows at once a simple formula relating the flux $\phi$ linked with a loop to the flux $\phi'$ linked with the image of this loop, provided each of the loops under consideration is again equidistant from the origin 0. The flux through such a loop can be calculated as $\phi=\oint\mathbf{A}d\mathbf{s}$, the integral being taken around the loop. The ratio of corresponding contributions to these integrals is (as $\mathbf{A}'\|\mathbf{A}$ and $d\mathbf{s}'\|d\mathbf{s}$)

$$\frac{\mathbf{A}'\cdot d\mathbf{s}'}{\mathbf{A}\cdot d\mathbf{s}}=\frac{ON}{K}\cdot\frac{ON'}{ON}=\frac{K}{ON}=\frac{\phi'}{\phi}. \qquad (14)$$

This is again a relation in which only the position of the "point of observation" enters, and not the position of the "causes."

If the flux through the loop (say, loop 1) is due to the fact that another loop (say, loop 2) carries a current $I_2$, we are led to a transformation formula for the mutual inductance $M_{12}$ between two such loops. It is

$$\phi_1=M_{12}\cdot I_2 \quad\text{and}\quad \phi_1'=M_{12}'\cdot I_2'. \qquad (15)$$

Hence

$$\frac{M_{12}'}{M_{12}}=\frac{\phi_1'/I_2'}{\phi_1/I_2}=\frac{K\cdot K}{r_1\cdot r_2}=\left(\frac{r_1'\cdot r_2'}{r_1\cdot r_2}\right)^{\frac{1}{2}}. \qquad (16)$$

The last formula is easily tested if we consider the mutual inductance between two circular loops, the axes of which intersect at an angle $\Theta$. (This angle can of course be zero, in which case we have two parallel, coaxial loops.) The value of this mutual inductance is given by Maxwell[6] by an expression of the form

$$M=r_2\cdot\Sigma_nC_n\left(\frac{r_2}{r_1}\right)^n=r_1\Sigma_nC_n\left(\frac{r_1}{r_2}\right)^n, \qquad (17)$$

where $r_1$ and $r_2$ denote the distances of the loops from the point 0 of intersection of the axes, and where the coefficients $C_n$ depend on the angle $\Theta$ and on the angles $\alpha_1$, and $\alpha_2$ which the radii of the loops subtend at *O* (see Fig. 3).

When this configuration is inverted, with 0 as center of inversion, we have

$$r_1'=K^2\ r_1, \quad r_2'=K^2/r_2$$

while $\alpha_1$, $\alpha_2$, and $\Theta$ remain unchanged. Hence

$$\frac{M'}{M}=\frac{r_2'\Sigma C_n\left(\dfrac{r_2'}{r_1'}\right)^n}{r_1\Sigma C_n\left(\dfrac{r_1}{r_2}\right)^n}=\frac{K^2}{r_1\cdot r_2}\,\frac{\Sigma C_n\left(\dfrac{r_1}{r_2}\right)^n}{\Sigma C_n\left(\dfrac{r_1}{r_2}\right)^n}=\frac{K^2}{r_1r_2} \qquad (18)$$

as required by Eq. (16).

If the loops are coaxial, any point on the common axis can be chosen as center of inversion. This can be used to transform the general case into various special cases. For instance, if we chose as center one of the points at which the common axis penetrates the sphere common to both loops, the two transformed loops will be co-planar. Another special case would be the transformation into loops of equal diameter, which is achieved by using as center of inversion the point at which the axis penetrates a sphere around the vertex of the cone common to both loops, the radius of this sphere being the geometrical mean of the distances of the two loops from this vertex (see the second method of apply-

---

[6] J. C. Maxwell, *A Treatise on Electricity and Magnetism*, Vol. 2, Chapter 14.

ing inversion to the problem discussed in the next section). It is, however, easily seen (by application of the formula given in the caption to Fig. 2) that in any such transformation the ratio of the shortest and the longest distance between these two loops remains invariant, which means (see below) that such transformation will not aid in evaluating the elliptic integral occurring in the formula for the mutual inductance.

## II. APPLICATIONS TO THE CASE OF A HIGH FREQUENCY SPHERICAL SCREEN

The theory developed in Section I can be usefully applied when we have to consider the screening effect of spherical metal "cans" at high frequencies. If the conductivity of such a screen is infinite, it is impermeable to magnetic flux at any frequency. Finite conductivity allows penetration of the field, though with the usual materials of construction such penetration is at radio frequencies limited to "skin depth." We may then either neglect this deviation from the ideal case altogether or take it into account by a simple approximation.

There are two ways in which the method of inversion can be applied to this problem, of which we shall consider first the more general one. Suppose that we are given a current loop—circular or otherwise—which is equidistant from the center of a hypothetical sphere, and further that we know the magnetic field produced by this loop. If we invert then from this center as center of inversion so that the sphere changes into itself, the loop will change into its image and we can predict from the theorems of Section I the field produced by this image. In particular, it follows from Eq. (14) that any test loop which we might draw on the surface of the sphere and that was linked with a certain amount, $\phi$, of the flux of the original loop will be linked with an identical amount, $\phi$, of flux produced by the image loop. Hence, if we take the correct sign for the current in the image loop, the sum total of the flux through the test loop will be zero. Thus the field produced by the combined action of original and image loop will have the property that none of its flux crosses the boundary of our hypothetical sphere, and we could introduce in its place an impermeable material screen without making any change to the field either inside or outside its boundary. Conversely, it follows that the effect of introducing such a screen on the field produced by the original loop could be simulated by the introduction of the image loop.

This allows us to calculate directly the change of self-inductance caused by the introduction of such a screen. We have, taking into account Eq. (10),

$$\Delta L = -M\cdot\frac{I'}{I} = -M\cdot\frac{r_0}{a}, \tag{19}$$

where $M$ is the mutual inductance between the image and the original loop and $r_0$ the distance of the latter from the center of the sphere of radius $a$.

Similarly, the flux arising from the image loop causes a change in the mutual inductance between the original loop and any other loop (say, loop 2) inside the sphere of amount

$$\Delta M_{12} = -M_{12}'\cdot\frac{r_0}{a}, \tag{19a}$$

where $M_{12}'$ denotes the mutual inductance between the image loop and the second loop.

In the case of circular loops formula (19) is easily evaluated, for the two loops are then coaxial and parallel (Fig. 4), and their mutual inductance is given (references 6 and 7) by

$$M = 8\pi\left(\frac{bb'}{k}\right)^{\frac{1}{2}}(K-E), \tag{20}$$

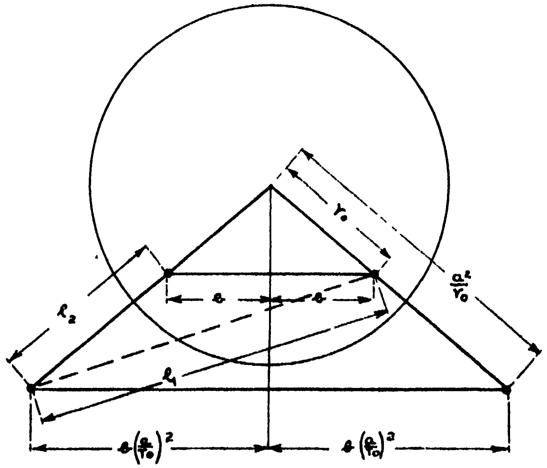

$r_0$
$\frac{a^2}{r_0}$
$e$
$e$
$\ell_2$
$\ell_1$
$e\left(\frac{a}{r_0}\right)^2$
$e\left(\frac{a}{r_0}\right)^2$

FIG. 4.

---

[7] A. Russel, J. Inst. Elec. Eng. 67, 659 (1929).

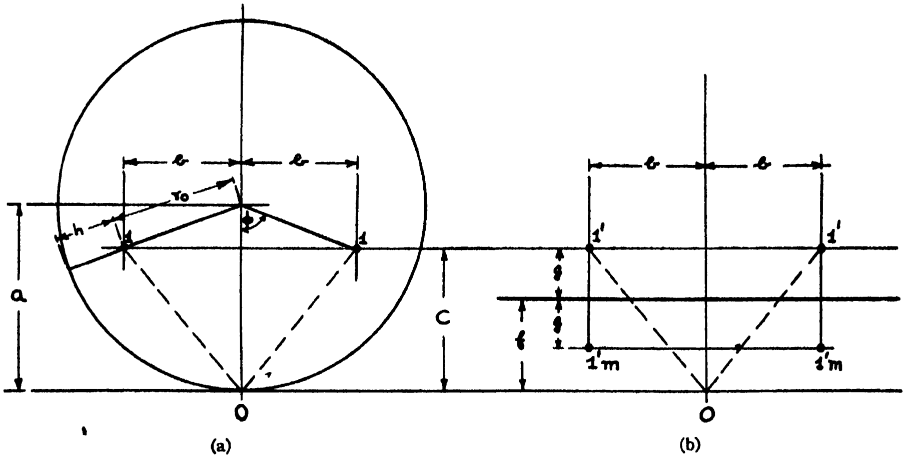


FIG. 5. Constant of inversion $K^2 = b^2 + c^2 = b^2 + (a - r_0 \cos\phi)^2$
$f = \frac{b^2+c^2}{2a}$ and $g = c - f = \frac{a^2 - r_0^2}{2a} \approx h$ for $h \ll a$.

where $b$ and $b'$ are the radii of the two loops and $K$ and $E$ are the complete elliptic integrals of the first and second kind to modulus $k$. This modulus itself is given by

$$k = \frac{l_1 - l_2}{l_1 + l_2} = \frac{\lambda - 1}{\lambda + 1} \left(\lambda = \frac{l_1}{l_2}\right), \tag{21}$$

where $l_1$ and $l_2$ denote the largest and the smallest distance between the elements of the two loops. In the present case $b' = ba^2/r_0^2$, and therefore

$$\Delta L = -\frac{8\pi b}{(k)^{\frac{1}{2}}}(K - E). \tag{22}$$

For the same reason we have, with $\alpha = a/r_0$, $\beta = b/r_0$ and $\nu = \alpha - 1/\alpha$

$$\lambda = \left(1 + \left(\frac{2\beta}{\nu}\right)^2\right)^{\frac{1}{2}}. \tag{23}$$

Evaluation of Eq. (20) is facilitated by extensive tabulation (references 6, 8, 9). There exists also a truly embarrassing number of approximations, (reference 9) so that it is perhaps useful to quote two which are simple but sufficient for slide-rule accuracy, (reference 7).

If $k < 0.9$, that is $\lambda < 20$ (that is in our case for loops which are not too close to the wall of the screen $P$) we have with an error of less than 6/10 of a percent

$$K - E = \frac{\pi k^2}{1 + 3(1 - k^2)^{\frac{1}{2}}}. \tag{24}$$

If $k$ is larger than 0.9 we have, with an error of less than ½ percent,

$$K - E = \tfrac{1}{2}(k)^{\frac{1}{2}}\left[\log_e 4 \cdot \frac{1+k}{1-k} - 2\right]. \tag{25}$$

The case of a very small loop can of course be dealt with in an elementary way. If the loop should be in the center of the sphere (Kaden's problem) we have the specially simple case that the image loop is infinitely distant, producing an image field that is homogeneous. The strength of this field can be found directly from the condition that the combined flux across the sphere must be zero everywhere. As the loop creates there a field strength (for unit current) equal to $2b^2\pi/a^3$ at those places where the sphere is penetrated by the axis of the loop, this too must be the strength of the image field. The image field sends then

$$\phi = b^2\pi \cdot 2b^2\pi/a^3 \tag{26}$$

---

[8] Circular C74 of the Nat. Bur. Stand. Washington (1924), reprinted 1937. A similar table is contained in Terman's *Handbook for Radio Engineers*. Further and more extensive tables are given by F. W. Grover, "Tables for the calculation of the mutual inductance of circuits with circular symmetry about a common axis," Sci. Pap. Bur. Stand. 20, (1926) and F. W. Grover *Inductance-Calculation* (D. Van Nostrand Company, Inc., New York, 1946).

[9] See the survey given by F. W. Grover in "Methods for the derivation and expansion for formulas for the mutual inductance of coaxial circles and of the inductance of single layer solenoids," Bur. Stand. J. Research 1, 487 (1928).

flux lines through the original loop, i.e.,

$$\Delta L = -2\pi^2 b\left(\frac{b}{a}\right)^3. \tag{27}$$

If the loop is placed eccentrically (distance $r_0$ from the center of the sphere) it follows similarly from Eq. (7) that

$$\Delta L = -2\pi^2 b\left(\frac{b}{a}\right)^3\left(\frac{\alpha^2}{\alpha^2-1}\right)^3. \tag{28}$$

The limiting case at the other end of the range is a loop that has from the screen only the very small clearance distance $h$. It can be shown that then the effect of the screen is identical with that of an impermeable plane that is parallel to the loop at a distance $h$. This particular case gives us an opportunity to illustrate shortly the second method of applying inversion which we mentioned at the beginning. Here we take as the center of inversion one of the points at which the axis of the loop penetrates the sphere. The sphere transforms then into a plane impermeable to flux. Without restricting the generality of the treatment we can select the constant $K$ of inversion so as to have the original loop transform into itself. With this particular choice of $K$ the current through the transformed loop will be of the same magnitude as in the original loop. The transformation is illustrated in Figs. 5a and 5b; this figure shows too how the transformed arrangement can be solved by the introduction of the mirror image $l_m'$ of the transformed loop $l'$.

If the walls of the spherical screen are not of infinite conductivity, the high frequency field penetrates to skin thickness and causes losses. These losses are calculated by Davidson and Simmonds for the case of the centrally situated loop. They can be calculated in a more general way if we make use of Wheeler's concept of the equivalent magnetic skin thickness ($=\frac{1}{2}$ conducting skin thickness). If the walls of the container of infinite conductivity are receded by the equivalent magnetic skin thickness, the magnetic energy of the field is increased (for constant input current) by just the same amount as if the field had penetrated into the walls of finite conductivity. (This is the definition of equivalent magnetic skin thickness.) The increase in loop reactance caused by this increase in field energy can easily be calculated, and, as the phase angle of such surface dissipation is always 45 degrees, this increase in reactance is numerically equal to the ohmic impedance component caused by the field penetration (reference 10).

In the present case, if the radius of the screen $a$ is increased by $\delta a$, the (negative) inductance increment $\Delta L$ previously found is varied by

$$\delta L = \left(\frac{\partial L}{\partial a}\right)\delta a. \tag{29}$$

Hence,

$$\delta R = 2\pi f\left(\frac{\partial L}{\partial a}\right)t_m, \tag{30}$$

where $f$ frequency of the field and $t_m = t/2$ equivalent magnetic skin thickness. The latter can be expressed in terms of the surface resistivity $R_s$ of the wall material. We have for the conducting skin thickness $t = 1/2\pi(\rho/f)^{\frac{1}{2}}$[$\rho$ = volume resistivity], and

$$t_m = \frac{t}{2} = \frac{R_s}{8\pi^2 f}$$

hence

$$t^2 = \frac{1}{4\pi^2}\frac{\rho}{f} = \frac{1}{4\pi^2}\cdot R_s\cdot t. \tag{31}$$

Equation (30) allows us to write down immediately Kaden's result for the resistance component. By differentiation of Eq. (27) we obtain

$$\delta R = \tfrac{3}{2}\pi\cdot\left(\frac{b}{a}\right)^4\cdot R_s.$$

There is no difficulty involved in evaluating Eq. (30) in a more general case, and, for instance, Davidson and Simmonds' result for $\delta R$ can be obtained from their series for $\delta L$ by term differentiation. Differentiation of the general expression given in Eq. (22) leads to a rather laborious expression which is therefore not given, as for all practical purposes it is sufficient to replace the differential quotient by a difference quotient, the latter being easily obtained by direct numerical calculation.

---

[10] H. A. Wheeler, Proc. I.R.E. **30**, 412 (1942). See also A. Bloch, W. Eng. **21**, 367 (1944).

# Non-Linear Viscous Elasticity and the Eyring Shear Model*

GEORGE HALSEY**

(Received April 17, 1947)

Attempts to analyze stress-strain-time relationships observed in experiments with textile fibers have been reported in a series of papers. This work has led to some general conclusions: First, the simplest three element model (spring in parallel with another spring and a dashpot in series) is adequate for the approximate explanation of a wide range of the data obtained. Second, the viscous element (dashpot) almost always appears to be non-Newtonian while the springs less frequently are non-Hookean. Third, the most striking deviations from simple behavior indicate, not a more complex model consisting of many simple elements, but that the assumed laws governing the viscous, or less frequently, elastic elements are themselves too simple.

Frequently a more complex law or laws will restore the three-element model to approximate agreement with the experimental facts. Finally, any system so far investigated can be made to show the presence of a number or a distribution of relaxation times by prolonging the experiment over a period of time, or more positively, by going to a point of no instantaneous relaxation, and waiting for the inevitable long period of relaxation to appear. Procedures to deal with the general situation described here are presented.

The paper begins with a condensed but essentially self-contained summary of the theoretical portions of the series of papers that preceded it. The original presentation has been clarified and recast in a form more in harmony with the present work. For experimental evidence, and supplementary information, the reader is referred to the original series.

The paper continues with a discussion of the various aspects of non-linear viscous elasticity, including the topics: Six rules for analyzing stress-strain curves (presented at the meeting of the Division of High-Polymer Physics of the American Physical Society, January 31, 1947); the thixotropic viscous element, where for convenience, complex situations, in which the populations of various states enter into a time-variable viscosity, have been gathered under the name "thixotropy"; constant force springs, where a particularly non-Hookean behavior is described.

## INTRODUCTION

TO make a beginning in understanding the stress-strain behavior of a material[1] like a textile fiber or a plastic, a simple mechanical analogy has much to recommend it. The stress-strain relationship of most materials is not independent of time; thus, a simple elastic element (a spring) does not suffice to duplicate the behavior. Since immediate elongation is generally observed upon increasing the load, however, a spring effect initially unrestrained by viscous forces at the rate of loading in question is required. The existence of hysteresis, which is almost never absent, indicates that a viscous element (dashpot) must be incorporated in the model. The viscous element is restrained from free motion, by a spring, because in a majority of cases, the flow at constant load does not continue at a constant rate, showing that the force initially concentrated on the viscous element ultimately shifts to a spring.

The arrangement of elements indicated by this qualitative analysis is shown in Fig. 1. The two arrangements are almost equivalent, and, when Hooke's law is written for both springs, they are represented by identical equations. When a model of this kind is elongated at a constant rate, the stress-strain curve will begin at a steep slope, and then bend over as the yield force of the dashpot is reached. In terms of the first arrangement of the three elements (Fig. 1*a*), the stress-strain curve is made up of a line of slope equivalent to the sum of the stiffnesses of both springs being elongated together, followed by a transition region, leading to a line of slope corresponding to

---

* Contribution of the Frick Chemical Laboratory, Princeton University and The Textile Foundation, Princeton, New Jersey

** Fellow of the Textile Research Institute.

[1] "Mechanical Properties of Textiles:" I. George Halsey, Howard J. White, Jr., and Henry Eyring, Textile Research J. **15**, 295 (1945). II. George Halsey and Henry Eyring, Textile Research J. **15**, 451 (1945). III. Henry Eyring and George Halsey, Textile Research J. **16**, 13 (1946), supplementary data: *ibid.* **16**, 284 (1946). IV. Richard Stein, George Halsey, and Henry Eyring, Textile Research J. **16**, 53 (1946). V. Henry Eyring and George Halsey, Textile Research J. **16**, 124 (1946). VI. H. D. Holland, George Halsey, and Henry Eyring, Textile Research J. **16**, 201 (1946). VII. George Halsey and Henry Eyring, Textile Research J. **16**, 329 (1946). VIII. Henry Eyring and George Halsey, Testile Research J. **16**, 335 (1946). IX. Sidney Katz, George Halsey, and Henry Eyring, Textile Research J. **16**, 378 (1946). X. C. H. Reichardt, George Halsey, and Henry Eyring, Textile Research J. **16**, 382 (1946). XI. C. H. Reichardt, and Henry Eyring, Textile Research J. **16**, 635 (1946). XII. C. H. Reichardt and George Halsey, Textile Research J. **17**, (1947) in preparation.

the stiffness of the spring in parallel with the dashpot (Fig. 2).

These relationships will be expressed mathematically. If the total force on the model is $F$, and its elongation (change in length/original length) is $l$, the condition of experiment is $l=\rho t$, where $t$ is the time and $\rho$ is the rate of elongation. The problem is to find $F$ as a function of $t$. If $f$ is the force on the dashpot, and $l_1$ its elongation, $F-f$ is the force on the spring in parallel, and $l-l_1$ is the elongation of the spring in series. Hooke's law will be assumed to hold for the springs, and Newton's law of viscosity will be written for the dashpot. Thus the problem being solved is the linear three-element model operating at a constant rate of elongation. The equations governing these elements are

$$F-f=k_2l, \qquad (1)$$

$$f=k_1(l-l_1) \qquad (2)$$

for the two springs, and

$$dl_1/dt=f/\eta \qquad (3)$$

for the dashpot, and

$$dl/dt=\rho, \qquad (4)$$

the condition of experiment.

At this point, it is to be noted that the force $F$ on the whole system can be separated into two parts easily recognizable on the stress-strain curve, shown in Fig. 2. The broken line in this figure represents a line parallel to the ultimate slope, and starting at the origin. This line gives the force, $F-f$, on the spring in parallel with the dashpot as a function of strain. $F-f$ is a very simple function of elongation alone, and attention can now be directed to the residue of the force on the system, the difference between the ordinates of this line and those of the stress-strain curve itself. This difference, of course, is $f$, the force on the dashpot and spring in series, making up a Maxwell element. This function of $f$ *versus* $t$ is referred to as the "working curve." Eliminating $l$ and $l_1$ from Eqs. 2–4 gives

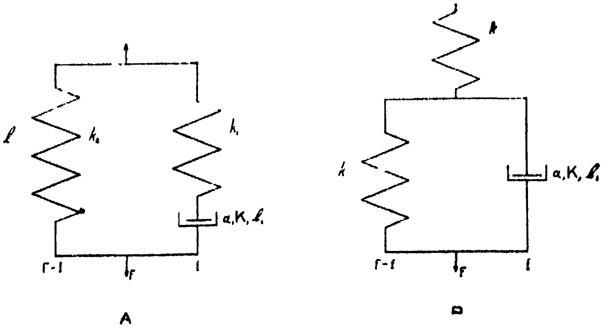


FIG. 1. Three-element model.

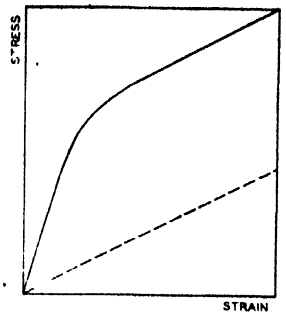


FIG. 2. A typical stress-strain curve.

$$(1/k_1\rho)(df/dt)=1-(f/\eta\rho). \qquad (5)$$

Writing

$$f/\eta\rho=\phi \qquad (6)$$

and

$$tk_1/\eta=\tau \qquad (7)$$

to absorb constants gives the dimensionless equation

$$d\phi/d\tau=1-\phi \qquad (8)$$

which can be integrated to

$$1-\phi=\text{const.}e^{-\tau}.$$

If $\phi=0$ when $\tau=0$ the constant becomes unity, and

$$\phi=1-e^{-\tau}. \qquad (9)$$

Now at this time it becomes desirable to introduce the idea of reduced variables. Equation (9) can be written in the form

$$af=1-e^{-bt}, \qquad (10)$$

where $a$ and $b$ are constants that must be adjusted to whatever units are selected for $f$ and $t$. If some natural unit of the curve is selected as the unit of measurement, what is called a reduced variable is defined. A procedure of this kind is often adopted when working with equations of state and it will be used here. Thus, the ultimate value of $f$ will be called unity, and $t$ will be called unity when $f=\frac{1}{2}$, in the reduced units. The constants $a$ and $b$ can then be com-

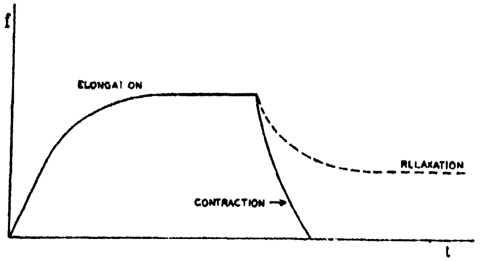


FIG. 3. Force on the dashpot as a function of time for elongation, contraction, and relaxation

puted when

$$t=\infty, \quad f=1, \qquad (11)$$
$$f=\tfrac{1}{2}, \quad t=1.$$

Therefore, from (10), $a=1$ and $\frac{1}{2}=1-e^{-b}$, so

$$b=\log_e 2=\ln 2$$

In reduced units, then,

$$f=1-e^{-t\ln 2}. \qquad (12)$$

The general nature of the curve is that of an exponential curve asymptotic to $f=1$, and for future reference, it should be noted that the force at the point $t=2$

$$f_2=1-\tfrac{1}{4}=0.75.$$

When the linear three-element model is relaxing at constant length, the problem is a very simple one, because $F-f$, the force on the parallel spring, is constant. Solving Eqs. (2) and (3) with $l$ constant gives

$$(1/k_1)(df/dt)=-f/\eta \qquad (13)$$

which yields a simple unimolecular decay law

The situation grows somewhat more complex when the usual repeated cycle stress-strain curve is considered. The separation of $F$ into $F-f$ and $f$ now has great utility. The force on the parallel spring is simply a function of elongation, and attention can be focused on the behavior of the dashpot and spring in series. Consider the working curve (Fig. 3) taken at a constant rate of elongation. If suddenly the elongation be halted, the relaxation curve begins, following Eq. (13). If the working curve be quite near its ultimate value, the dashpot is flowing at a rate equal to $\rho$, the over-all rate of elongation and the spring has ceased to elongate. This is shown by the constant value of the force. Therefore, when the elongation ceases, the dashpot will be flowing on at a rate equal to the over-all rate, and the force on the series spring will decay at exactly the same instantaneous rate as the initial slope near $t=0$ when the spring was being loaded at the over-all rate, $\rho$.

Now, if instead of stopping for relaxation the rate of elongation is reversed, at the ultimate value of $f$, then the force will be decaying for two reasons. The dashpot is still elongating at an instantaneous rate $\rho$, and the series spring is being shortened by another $\rho$. The total rate of decay is $2\rho$, and so, the instantaneous slope upon reversal when $f_{\text{reduced}}=1$ is twice the initial slope. As the contraction at rate $\rho$ proceeds, the slope decreases, for obviously, when $f=0$ the slope must be the initial slope, because the dashpot is no longer moving at all. As further contraction into the region of negative tension occurs, the situation is the same as in the initial elongation except that the sign of $f$ becomes negative.

The equation for this region is still the familiar (12). Then, because the region of steep slope occurs without an intervening reversal at a time preceding the familiar region, it becomes apparent that the equation for this region is also (12), but in the region of negative time. The slope at any value of $\phi$ is obtained from (8), and it is seen that at $\phi=-1$, the slope is twice that of $\phi=0$.

The working curve for positive and negative values of time is shown in Fig. 4. Using this curve, the stress-strain curve for many cycles of elongation and contraction is easily constructed with a pair of dividers and a ruler. The sawtooth pattern of the force on the spring in parallel is first marked out, putting in a horizontal sec-

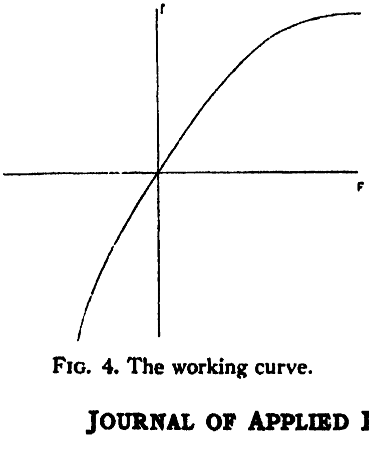


FIG. 4. The working curve.

tion for every relaxation. Then, starting at any point when $f$ is known to be zero, the curve is laid out measuring the $f$ value from the working curve. Whenever the direction of elongation reverses, the force value is measured from the other branch of the working curve, that is, where the value of $f$ is the same except for sign. Now it is quite clear that if the shape of the working curve is known, this process of calculation of a complex curve can be generalized. None of the laws has to be linear. The only restriction, at the moment, is that the viscous drag at a given rate be the same for moving the dashpot forward as for moving it backward. Actually, to calculate the working curve depends on the ability to integrate the equation, but this integration can always be found by numerical means.

With this generality available, the question arises "when is the three-element arrangement itself permissible?" The answer to this question lies in a study of the points where $f=0$, that is, when the stress-strain curve crosses the line ($F-f$ *vs.* $l$) representing the force on the spring in parallel. At this point, there will be no relaxation. On either side of it, relaxation will occur in opposite directions, tending toward it. These "points of no relaxation" lie on a line that is a function of elongation alone. By a study of experimental points of no relaxation, it can be ascertained that they are a function of elongation, alone. If they are, their totality of points forms the line representing $F-f$, the force on the parallel spring, which for convenience is referred to as the "spring line."

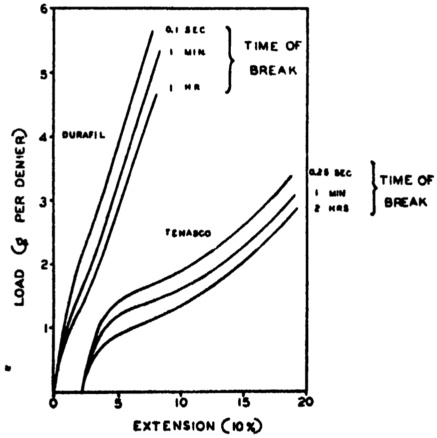


FIG. 5. Effect of rate of loading on load-elongation properties of regenerated cellulose rayons at 20°C and 65 percent R.H.

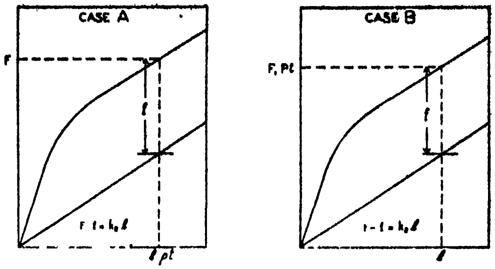


FIG. 6. The manner in which the various quantities are determined from the stress-strain curve; *A*, at constant rate of elongation, and *B*, at constant rate of loading.

A change in rate of experiment produces quite striking results. On the assumption of the Newtonian viscous element the force at a given elongation should be a linear function of the rate of elongation. Contrary to this, the function is found to be more nearly logarithmic. For example, D. Entwistle, in the third John Mercer Lecture,[2] publishes a curve (Fig. 5) showing the rate of elongation changed over a factor of 28,800 with the curves arranged in such a manner that he states: "Over a considerable range there is an approximately linear relationship between breaking load and the logarithm of the time of break." Because the breaking elongation is relatively independent of rate, this is equivalent to stating that the logarithmic law holds at any given elongation, since the curves parallel one another. This situation is entirely incompatible with any linear model, but harmonizes very well with the Eyring viscous element, or for that matter, any exponential law of flow.

In 1936 Eyring[3] proposed his law for viscosity in shear:

$$\text{rate of shear} = 2(V_h/V_m)(kT/h)e^{-\Delta F\ddagger/kT} \sinh f V_h/2kT, \quad (14)$$

where $V_h$ is the volume of the flow "hole," $V_m$ is the volume of the molecule, $k$ is Boltzmann's constant, $T$ is temperature, $h$ is Planck's constant, $\Delta F\ddagger$ is the free energy of activation for flow, and $f$ is the shear stress. This can be written for the present as

$$dl_1/dt = K \sinh \alpha f. \quad (15)$$

---

[2] D. Entwistle, J. Soc. Dyers and Colourists 62, 261 (1946).

[3] Henry Eyring, J. Chem. Phys. 4, 263 (1936).

The problem of the three-element model incorporating this viscous element will now be solved. It is a quite general law, and it includes the Newtonian law as a special case. The equations governing the model are:

$$F-f=k_2l,$$
$$f=k_1(l_1-l_1),$$
$$dl_1/dt=K\sinh\alpha f,$$

and, for the condition of experiment,

$$pF+ql+rt=0.$$

This equation represents the most general restraint that can be integrated conveniently. It includes, as special cases, a constant rate of elongation, $\rho$, and a constant rate of loading, $P$. For these two cases the various variables are measured in the manner shown in Fig. 6. Also governed by this equation are the results of the pendulum-type tester, if the movement of the upper jaw is proportional to the load. Solving these equations for $f$ as a function of $t$ yields the equation

$$-\left(\frac{1}{k_1}+\frac{p}{pk_2+q}\right)\frac{df}{dt}-\frac{r}{pk_2+q}=K\sinh\alpha f. \quad (16)$$

This equation is of the form

$$\beta-d\phi/d\tau=\sinh\phi, \quad (17)$$

which is integrated to give

$$\frac{2}{(1+\beta^2)^{\frac{1}{2}}}\tanh^{-1}\left\{\frac{e^{+\phi}-\beta}{(1+\beta^2)^{\frac{1}{2}}}\right\}=\tau-c$$

Solving this equation for $\phi$ gives

$$\phi=\ln\left\{\beta+(1+\beta^2)^{\frac{1}{2}}\tanh\left[\frac{(1+\beta^2)^{\frac{1}{2}}}{2}(\tau-c)\right]\right\}. \quad (18)$$

When $\tau=0$, $\phi=0$, so

$$c=\frac{2}{(1+\beta^2)^{\frac{1}{2}}}\tanh^{-1}\left\{\frac{\beta}{(1+\beta^2)^{\frac{1}{2}}}\right\}. \quad (19)$$

Now, it is desired to obtain a reduced equation, so following a procedure like that employed in the linear case, the equation

$$af=\ln\left\{\beta+(1+\beta^2)^{\frac{1}{2}}\tanh\left[\frac{(1+\beta^2)^{\frac{1}{2}}}{2}(bt-c)\right]\right\} \quad (20)$$

is solved for $a$ and $b$, under the conditions that when $t=\infty$, $f=1$, and when $t=1$, $f=\frac{1}{2}$. After a series of algebraic manipulations the reduced equation

$$f=\frac{\ln\left\{\dfrac{1+\dfrac{1+\beta}{(1+\beta^2)^{\frac{1}{2}}}\tanh(At)}{1+\dfrac{1-\beta}{(1+\beta^2)^{\frac{1}{2}}}\tanh(At)}\right\}}{\ln\left\{\dfrac{1+\dfrac{1+\beta}{(1+\beta^2)^{\frac{1}{2}}}}{1+\dfrac{1-\beta}{(1+\beta^2)^{\frac{1}{2}}}}\right\}} \quad (21)$$

is obtained, where

$$A=\tanh^{-1}\left\{\frac{\left[\left(\dfrac{1+\beta}{(1+\beta^2)^{\frac{1}{2}}}+1\right)\left(\dfrac{1-\beta}{(1+\beta^2)^{\frac{1}{2}}}+1\right)\right]^{\frac{1}{2}}-1}{\left(\dfrac{1+\beta}{(1+\beta^2)^{\frac{1}{2}}}+1\right)\left(\dfrac{1-\beta}{(1+\beta^2)^{\frac{1}{2}}}+1\right)-1}\right\}. \quad (22)$$

With this function of $t$ and $\beta$ it is possible to calculate a whole family of working curves, depending on the shape factor, $\beta$. The larger $\beta$ is, the more non-Newtonian is the behavior of the working curve. To compute $\beta$ from the working curve, the behavior of the reduced force at some suitable value of $t_{\text{red}}$ can be investigated. It was shown above that for the Newtonian case for $t=2$, $f=0.75$. In the non-Newtonian case this force is always larger, approaching unity as $\beta$ increases indefinitely. The quantity $f_2$ *versus* $\beta$ is shown in Fig. 7.

The meaning of the increase of $f_2$ with $\beta$ is found in the fact that the more non-Newtonian a dashpot is, the more insensitive it is to forces below the yield force. In the limit, the com-

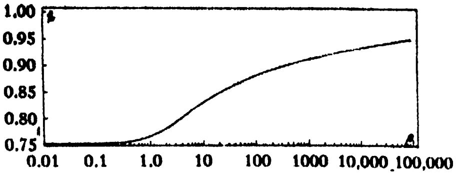


FIG. 7. Diagram for evaluating $\beta$ from the reduced force, $f_2$, acting on the dashpot at a time, $t_2$, which is twice the time elapsed from the application of the load until it reaches one-half its final value.

pletely non-Newtonian dashpot yields quite sharply at a given force, independent of the rate of elongation, and does not move below this force. For practical purposes of constant evaluation the shape factor is only of the roughest significance, but at this point it is appropriate to give the constants of the viscous law as a function of the observed quantities. For the case of a constant rate of elongation $K=\rho/\beta$ and $\alpha=1/f_\infty$ arc sinh$\beta$. The ultimate value of the force on the dashpot is written $f_\infty$. Similarly for the case of constant rate of loading, $P$, $K=P/\beta k^2$ and

$$\alpha=1/f_\infty \text{ arc sinh}\beta.$$

Various other ways of evaluating $\beta$ are given in the original references, along with a treatment of relaxation and creep and a more detailed discussion of the Eyring viscous element.

The practicable way of evaluating $\beta$ is found when the change of the ultimate force, $f_\infty$, with rate of elongation is considered. If $r_n$ is the ratio of $f_\infty$ at a speed $n$ times a basic speed $\beta$, to $f_\infty$ at the unit speed, then one has

$$r_n=f_\infty(n)/f_\infty(1)=\phi_n/\phi_1, \tag{23}$$

$$\phi_1=\text{arc sinh}\beta,$$
$$\phi_n=\text{arc sinh}n\beta,$$

and, therefore,

$$r_n=\text{arc sinh}n\beta/\text{arc sinh}\beta. \tag{24}$$

Therefore, $\beta$ can be determined as a function of $r_n$. If $\beta$ is large, the approximation arc sinh$\beta$ $\approx \ln 2\beta$ holds, and $\beta$ can be found explicitly:

$$\beta=\tfrac{1}{2}n^{(1/r_n-1)}. \tag{25}$$

This equation provides a way to determine $\beta$ as a function of scale factors alone, and independent of shape.

The three-element non-Newtonian model is quite successful in accounting, qualitatively at least, for many stress-strain curves. It should be emphasized that the non-Newtonian nature of the dashpot is responsible for its remarkable qualitative fidelity. In complicating the three elements, very little progress can be made merely by having many Newtonian elements hooked together. Most existing material concerning the visco-elastic behavior of materials, however general and formal, has in it somewhere the assumption that the governing equations are linear. This is particularly true of the so-called superposition principle. The utility of such schemes is thus strictly limited. In fact, linear systems are usually found only in highly plasticized systems where the stresses are in the region of $10^7$ dynes per square centimeter. In textiles, where stresses are of the order of $10^9$ dynes per square centimeter, no case has come to the author's attention. Further discussion of this topic is found in the original references. The most fruitful complications are found when the elements themselves are changed, retaining the three element model.

In deriving the Eyring viscous law, the assumption is made that flow is retarded by a symmetrical potential barrier. It was always realized that this simplification might be unjustified, and it has been observed that it is often not available. That is, the force, notably in the case of wool, is much more effective in producing flow in one direction than in the opposite one. Clearly, if the potential barrier is unsymmetrical, the rate of reaction up the steep side will be less affected by the force than the rate up a more gradual slope. This, then, is in harmony with the experimental observation. With an unsymmetrical barrier, the expression for the rate of shear is

$$dl_1/dt=(K/2)(e^{2\mu\alpha f}-e^{-2(1-\mu)\alpha f}), \tag{26}$$

where $\mu$ is the symmetry coefficient. The form

$$dl_1/dt=Ke^{+(2\mu-1)\alpha f}\sinh\alpha f \tag{27}$$

is equivalent to this expression, which shows the perturbing influence of the lack of symmetry on the original hyperbolic sine law. For convenience, $\sigma$ is written for $2\mu-1$.

Using this new viscous law, the equation governing the working curve becomes

$$\beta-(d\phi/d\tau)=e^{+\sigma\phi}\sinh\phi, \tag{28}$$

which gives

$$-\tau=\int\frac{d\phi}{\beta+e^{\sigma\phi}\sinh\phi}, \tag{29}$$

which is not integrated. The measurement of the ratio

$$(f_n/f_1)_\infty=(\phi_n/\phi_1)_\infty$$

still permits determination of the constants, even if the exact shape of the working curve is not calculated.

Using the notation $\phi_1{}^{\text{out}}$ for the ultimate value of $\phi$ with the dashpot going forward at speed 1, and similarly $\phi_n{}^{\text{out}}$, $\phi_1{}^{\text{in}}$, and $\phi_n{}^{\text{in}}$ for the other three ultimate values, the equations when $d\phi/d\tau$ becomes zero, are

$$\begin{aligned}\beta&=e^{+\sigma\phi_1{}^{\text{out}}}\sinh\phi_1{}^{\text{out}},\\ n\beta&=e^{+\sigma\phi_n{}^{\text{out}}}\sinh\phi_n{}^{\text{out}},\\ \beta&=e^{-\sigma\phi_1{}^{\text{in}}}\sinh\phi_1{}^{\text{in}},\\ n\beta&=e^{-\sigma\phi_n{}^{\text{in}}}\sinh\phi_n{}^{\text{in}}.\end{aligned} \tag{30}$$

Using the approximation $\sinh x\approx e^x/2$ for large $x$, these equations become

$$\begin{aligned}\ln 2\beta&=(1+\sigma)\phi_1{}^{\text{out}},\\ \ln n+\ln 2\beta&=(1+\sigma)\phi_n{}^{\text{out}},\\ \ln 2\beta&=(1-\sigma)\phi_1{}^{\text{in}},\\ \ln n+\ln 2\beta&=(1-\sigma)\phi_n{}^{\text{in}}.\end{aligned} \tag{31}$$

From this set of equations, a set of ratios may be derived as follows:

$$R^{\text{out}}=(\ln 2\beta+\ln n)/\ln 2\beta=R^{\text{in}}, \tag{32}$$

$$R_n=(1+\sigma)/(1-\sigma)=\phi_n{}^{\text{in}}/\phi_n{}^{\text{out}}; \tag{33}$$

where

$$\begin{aligned}R^{\text{out}}&=\phi_n{}^{\text{out}}/\phi_1{}^{\text{out}}=f_n{}^{\text{out}}/f_1{}^{\text{out}},\\ R^{\text{in}}&=\phi_n{}^{\text{in}}/\phi_1{}^{\text{in}}=f_n{}^{\text{in}}/f_1{}^{\text{in}},\\ R_n&=f_n{}^{\text{in}}/f_n{}^{\text{out}}.\end{aligned}$$

From these equations for large $\phi$ $\beta=\frac{1}{2}n^{1/R-1}$ (as is found for this approximation when $\sigma=0$), and

$$\sigma=\frac{R_1-1}{R_1+1}=\frac{R_n-1}{R_n+1}, \tag{34}$$

or

$$\mu=\frac{R_1}{R_1+1}=\frac{R_n}{R_n+1}, \tag{35}$$

where the formulas with $R_n$ are obviously better approximations.

In the other approximation, where $\phi$ is small, the four equations become so simple that $\sigma$ disappears completely, in harmony with the necessary equality at the rates of reaction at equilibrium, satisfying the principle of detailed balancing.

As a second degree of approximation, consider that $\beta=\beta_0+\beta_1$, $\sigma=\sigma_0+\sigma_1$, and $\phi_1{}^{\text{out}}=\phi_0+\phi_1$, where $\phi_0=(1+\sigma_0)^{-1}\ln 2\beta_0$. Upon expansion in the small correction factors, using the approximate equation for the zero-order values, closer values for the constants can be found.

Of some utility is the limiting concept of the completely unsymmetrical barrier, where $\mu=1$. Flow in one direction is not affected by force at all. The equation for shear becomes

$$dl_1/dt=(K/2)(e^{2\alpha f}-1). \tag{36}$$

If the slightly altered definitions $\phi=2\alpha f$, $\beta=(2\rho/K)+1$, and

$$\tau=k_1\alpha Kt$$

are made, the differential equation for the working curve becomes

$$\beta-(d\phi/d\tau)=e^{\phi}, \tag{37}$$

which yields

$$\tau+\text{const.}=(1/\beta)[\phi-\ln(\beta-e^{\phi})]. \tag{38}$$

A reduced equation will be obtained, subject to the conditions, that when

$$\begin{aligned}t&=0, & f&=0\\ t&=\infty, & f&=1\\ t&=1, & f&=\tfrac{1}{2}.\end{aligned}$$

The equation is

$$t=\frac{(f\ln\beta-\ln(\beta-e^{f\ln\beta})+\ln(\beta-1))}{(\ln\sqrt{\beta}-\ln(\beta-\sqrt{\beta})+\ln(\beta-1))}, \tag{39}$$

an equation which changes when $-f$ is substituted for $f$.

The only type of experiment considered above was one in which a linear relationship existed between force, elongation, and time:

$$pF+ql+rt=0,$$

because only this restraint makes the equations amenable to simple treatment. In practice, it is often desirable to calculate behavior that is characterized by its dependence on any restraint, assuming in this discussion that the constants of the model have been evaluated from simple experiments and hence are available for computation. However, in solving even the simplest cases analytical difficulties are met, and it is of interest to consider here simple point-by-point methods. The method of synthesizing the total force, $F$, from the partial forces, $f$, on the dashpot and spring in series, and $F-f$, on the single spring will be employed, as before. If the restraint is of the form $l=R(t, F)$, then $F-f=k_2l=k_2R(t, F)$. The value of $f$ must be considered with more care. First the case of constant rate is considered. The value of $\phi$ is

$$\phi=\ln\left\{\frac{1+\dfrac{1+\beta}{(1+\beta^2)^{\frac{1}{2}}}\tanh\dfrac{\tau}{2}(1+\beta^2)^{\frac{1}{2}}}{1+\dfrac{1-\beta}{(1+\beta^2)^{\frac{1}{2}}}\tanh\dfrac{\tau}{2}(1+\beta^2)^{\frac{1}{2}}}\right\} \tag{40}$$

with a new set of normalization conditions.

The time values were normalized previously to the half-value of the ultimate force (that is, when $t=1, f=\frac{1}{2}$), but now the ratios between the initial slopes must be preserved in their proper values. This is made necessary by the new variable that is now considered. Thus, the three conditions defining the old reduced variables variables were as follows: when

$$\begin{aligned} f&=0, & t&=0,\\ f&=\tfrac{1}{2}, & t&=1,\\ f&=1, & t&=\infty, \end{aligned}$$

and the definitions of the new reduced variables are as follows: when

$$\begin{gathered} f=0, \qquad t=0\\ Kak_1=1,\\ \alpha=1. \end{gathered}$$

This curve, or family of curves, depending on $\beta$, is adequate to predict the value of $f$ from its initial value, $f_0$, as long as $f_0$ is below $f_\infty$ in value. For values of $f_0$ greater than $f_\infty$ it is necessary to determine the family of curves where the extreme boundary value $f_0=\infty$ is employed. The general solution

$$f=1/\alpha\ \ln[\beta+(1+\beta^2)^{\frac{1}{2}}\tanh(bt+c)] \tag{41}$$

yields at $f=\infty$, $t=0$

$$\begin{aligned} \infty &=\ln[\beta+(1+\beta^2)^{\frac{1}{2}}\tanh c],\\ c&=\pi/2(-1)^{\frac{1}{2}}. \end{aligned}$$

This yields, as a special solution,

$$\phi'=\ln\left\{\beta+\frac{(1+\beta^2)^{\frac{1}{2}}}{\tanh bt}\right\}. \tag{42}$$

(The constant $b$ is then $(1+\beta^2)^{\frac{1}{2}}$ to agree in scale with $f$.) This equation represents a family of curves asymptotic to $t=0$ and $f'=f_\infty$, because when $t=\infty$

$$\phi_\infty'=\ln[\beta+(1+\beta^2)^{\frac{1}{2}}], \tag{43}$$

and for the other set of curves passing through $t=0, f=0$

$$\phi_\infty=\ln\left\{\frac{1+\dfrac{1+\beta}{(1+\beta^2)^{\frac{1}{2}}}}{1+\dfrac{1-\beta}{(1+\beta^2)^{\frac{1}{2}}}}\right\} \tag{44}$$

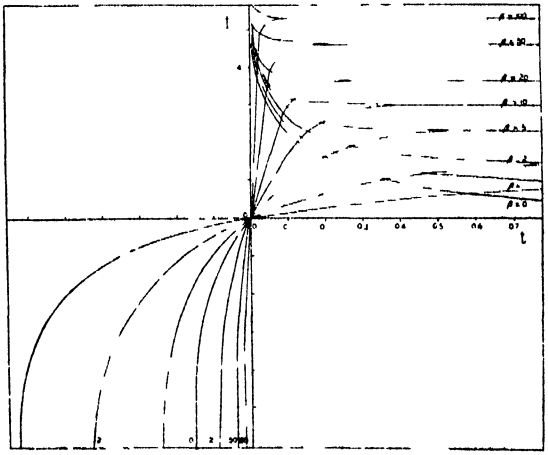


FIG. 8. The working curves for various values of $\beta$ from 1 to 100.

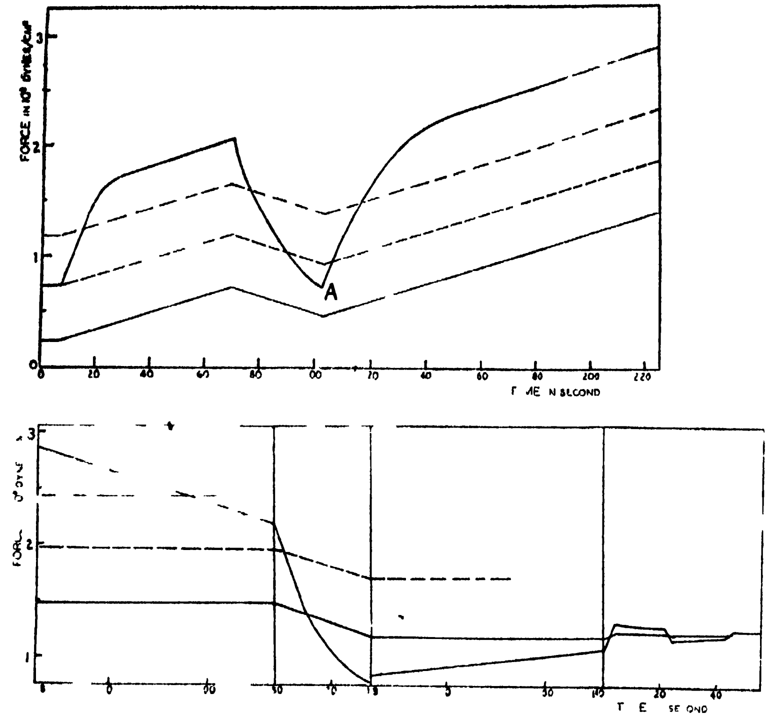


FIG. 9 A typical stress-strain curve showing the inadequacy of the three-element model It is 6-denier textile rayon, elongated at a rate $6\ 5\times 10^{-4}$ per second at 0 35 R H

and, since

$$\beta+(1+\beta^2)^{\frac{1}{2}}=\frac{1+\dfrac{1+\beta}{(1+\beta^2)^{\frac{1}{2}}}}{1+\dfrac{1-\beta}{(1+\beta^2)^{\frac{1}{2}}}}=e^{\text{arc sinh}\,\beta},$$

$$\phi_\infty{}'=\phi_\infty$$

These two families, which together cover the entire range of $f$ from $-\infty$ to $\infty$, are plotted in Fig. 8. Thus with a given velocity and initial conditions, subsequent conditions can be predicted at that velocity by following the proper curve But since

$$l=R(F,t),$$

or $$dl/dt=\rho=R'(F,t) \tag{45}$$

is no longer a constant, it is possible to predict the initial velocity if $f_0$, $F_0$ and the constants are known to be

$$\beta_0=\rho_0/K=R'(F_0,0)/K. \tag{46}$$

Proceeding along the proper curve $\beta_0$ an amount $\Delta t_1$, from $f_0$ gives

$$\Delta(F-f)_1=\rho_0 k_2\Delta t. \tag{47}$$

Here $k_2$ is the force constant of the parallel spring $\Delta F_1$ is obtained from Fig 8

$$\Delta l_1=\rho_0\Delta t_1$$

Then a new $\beta$ is determined

$$\beta_1=\rho_1/K=R'[(F_0+\rho_0 k_2\Delta t_1+\Delta f_1),\ \Delta t_1]/K \tag{48}$$

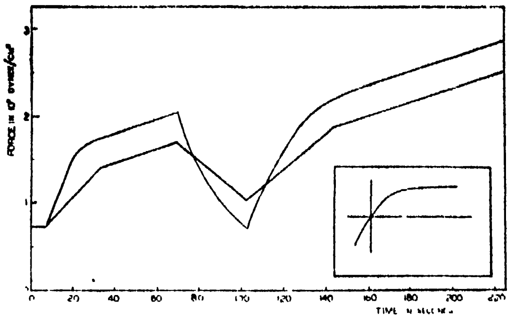
Then, from the point $f_0+\Delta f_1$, curve $\beta_1$ is followed for a time $\Delta t_2$, and another point is thus constructed. This process is repeated until the required curve is produced.

Now, in order to work a specific problem where the values of the constants are presented in centimeter-gram-second units, it is necessary to reduce the units. First, the value of the force unit is taken, so that $\alpha$ becomes unity in terms of the new force variable. Then, time is taken so that the product $K\alpha k_1$ equals unity, whereupon the variables are reduced. Then $\beta_{\text{initial}}$ is computed from $\rho_{\text{initial}}$ and $K$, and $k_2$ are expressed in terms of reduced force, and the computation begins. The only numbers that characterize a given computation, then, are the experimental restraints such as inertia (which must again be in reduced units), the value of $k_2$ in reduced units, and $\rho_{\text{initial}}$.

Although a large amount of material can be explained with the three-element model, almost any material can be made to show the presence of a distribution of relaxation times by a suitably prolonged experiment. Before going into the general theory of complex arrangements of non-linear elements, it will be helpful to consider a simple system that shows some of the qualitative features of the distribution.

The use of any sort of model containing more than one viscous element is a difficult undertaking. However, the most rudimentary considerations of the molecular structure of a fiber make it seem highly unlikely that the fiber can be described with a single viscous element, no matter how involved its law of operation. Nevertheless, a single element serves to a remarkable degree to reproduce the behavior of materials which do not exhibit plastic flow to an appreciable extent. The "distribution" used here contains only two elements; no particular significance, however, can be attached to this, except perhaps that the data do not warrant further extension. The actual fiber, no doubt, involves a continuous distribution. It is necessary to remark that this distribution has nothing in common with the Wiechert distribution of linear elements; this latter is totally inadequate to explain the experiment. It is much farther from the facts than the three-element, non-Newtonian model.

In the analysis of a large number of stress-strain curves, using the method of the three-element model of previous papers, certain persistent difficulties have been encountered. The curve of Fig. 9 shows an extreme case. The ultimate slope is well defined. Therefore, simple permanent set cannot be a factor over moderate time ranges. Repeated cycles could then be expected to yield a limiting curve. In order to place the line of ultimate relaxation, three procedures are resorted to, using the well-defined ultimate slope. First, the line is drawn from the origin. This is unsatisfactory for two reasons: the line does not cut the experimental curve with constant angle and it does not agree with the position of ultimate relaxation as determined at the end of the experiment. An attempt to draw a line through the points of constant slope meets with even less success. Using the established slope, it is impossible to go through all of the points. The line is too high for the relaxation curve, and, of course, it fails completely to agree with the determination of the point of ultimate relaxation. Projecting the line back from this point also fails completely. It leaves the point *A*, which is obviously a point below the spring line, hopelessly above it.



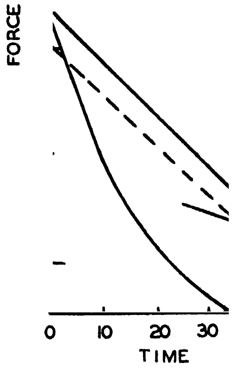


FIG. 11. The curve of Fig. 9 here is analyzed on the basis of the model of Fig. 10. Only a small excerpt from the second part of Fig. 9 is shown.

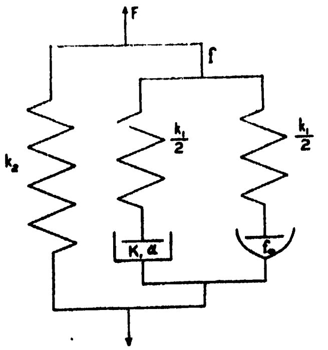


FIG. 10. The five-element model used in this paper to explain the curve of Fig. 9.

These considerations obviously lead to the conclusion that the model is not a correct picture of this fiber. However, the qualitative as-

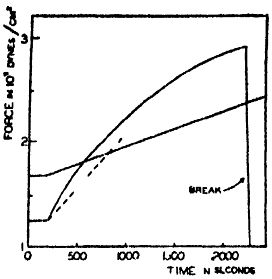


FIG. 12. The experiment of Fig. 9 carried out at one twenty-fifth the former rate. It shows the presence of a distribution but is interrupted by breakage of the fiber

pects of the diagram are the same as a curve produced by a perfect three-element model. No widely separated processes exist. The sharpness of the initial yield point completely disallows a wide distribution of yield forces of a simple non-thixotropic character. These yield forces must be within perhaps 10 percent of one another. Equally impossible is any type at all of three-element model. A single dashpot, even if unsymmetrical and thixotropic, cannot explain point *A*. That is, a true point of ultimate relaxation cannot always be achieved by finite process.

The model suggested by these considerations is one containing a number of elements of the same yield strength. A simple method of this type suffices to explain the curve in Fig. 9. This model (Fig. 10) contains two dashpots of equal yield value. One is quite non-Newtonian and the other is more nearly Newtonian. The effect of this arrangement is that the non-Newtonian dashpot reinforces the spring when it is not at the yield force. When the point of ultimate relaxation is found by hunting, it does not represent the place where the force on the compound element is zero, but the place where the force on the Newtonian dashpot is zero. Therefore, the slope to be used in leaving this point to connect with other similar points is not the well-defined slope but a steeper slope calculated by connecting several points of the proper slope. This slope must yield to the well-defined slope when the yield value is reached. That this model accounts for the curve is apparent (Fig. 11). During the relaxations, the model must be abandoned in favor of a smoother distribution of some unknown character. At a finite rate, however, this crude model is the most complicated that can be justified. Its importance is that it is the *simplest* model that possibly can give rise to the experimental facts.

When the force is divided among several viscous elements, there are at least two ways to express the constants. If the actual force on an element is used to compute the constant, then this represents the exact mechanical element that would reproduce the behavior when combined with its fellow springs and dashpots. However, if constants are based on the total viscous force along with a "concentration," then the situation is conceptually clearer. Thus, if one desires to make a three-element model with two identical dashpots instead of one, it seems desirable that the constants should be the same as if he had used only one because the behavior would be identical.

So far as a single rate of elongation is concerned, the model of Fig. 10 seems quite adequate. Figure 12 shows the effect of changing the rate of elongation to one twenty-fifth of its former value. Unfortunately, the viscose rayon fiber soon broke The shape of the curve, however, shows that the hypothesis of equal yield forces is no longer good The presence of a distribution of relaxation times is clearly indicated. This is as it should be, for if at one speed the yield force of the dashpots is equal, then at other speeds the yield forces must diverge from one another. Figure 12 shows that the dashpot of the more Newtonian character is almost completely yielded at all times, and that the yield force is approximately accounted for by the completely non-Newtonian dashpot if that yield force, by hypothesis, is considered constant. The gradual nature of the yielding process shows that this hypothesis is only qualitatively correct, and that the completely non-Newtonian dashpot is beginning to split into several more or less non-Newtonian dashpots.

Proceeding in a manner similar to that described above and observing the convention therein described, one obtains the following results:

Springs: $k_1 = 0.94 \times 10^{11}$ dyne/cm$^2$,
$k_2 = 1.19 \times 10^{10}$ dyne/cm$^2$;
First dashpot: 0.66 concentration
$f_\infty = 0.874 \times 10^9$ dyne/cm$^2$;

Second dashpot: 0.34 concentration

$\rho = 6.5 \times 10^{-4}$,

$\beta = 3$ (from shape),

$k = 2.2 \times 10^{-4}$,

$\alpha = 2.09 + 10^{-9}$,

$\Delta F\ddagger = 23{,}650$ calories,

$V_h = 5.5$ A cubed.

In Fig. 11, second part, the agreement with the established line of no relaxation is shown by the difference between the dashed line and the solid line. In Fig. 12, the dashed line gives the slope used before to represent the single dashpot. At this new speed, it no longer has meaning.

## SIX RULES FOR ANALYZING STRESS-STRAIN CURVES

In the analysis of the stress-strain properties of a material there is a certain advantage to the making of haphazard stress-strain curves, following no particular pattern and depending, almost, on the whim of the experimenter. Then these curves are subjected to an analysis that must meet many novel conditions of experiment. For path-finding purposes this procedure is very useful. Until the present writing, stress-strain work in the laboratory has followed such a pattern, or lack of pattern. However, as more painstaking analyses are attempted, and as the model is refined, this procedure becomes very wasteful and the results meager. An attempt has been made to improve the situation by setting up a series of six rules for the analysis of the stress-strain curves of a material.

In the present paper, the general distribution of non-Newtonian relaxation times is included in the scheme of analysis. The introduction of the distribution has been avoided in previous papers because its great complexity obscured the exposition of simpler ideas, such as non-Newtonian behavior and the unsymmetrical potential barrier. The importance of the distribution is realized when more complex experiments are performed. It is seldom if ever found absent when experiments are devised that accentuate it. Similarly, the existence of multiple equilibrium positions at the same value of the flow coordinate (thixotropy) is of very great importance. The protein fibers seem especially thixotropic. Because both of these complications are introduced here, the scheme is more general in application than the analysis given in previous papers.

Each of the proposed six rules contains a simplifying criterion. If the criterion in the first rule is satisfied, then the procedure of the second rule becomes usable, and so on through the rules. If all six criteria are obeyed, then the material under study is representable by a hyperbolic sine-law model. When a criterion is violated, the notes to the rules suggest the proper procedure. At times this procedure can only be sketched, because the mechanics of the complex analysis is still being worked out. However, the principles governing the calculation are usually clear. The six rules suggest an orderly mode of experiment which is outlined in a note. The language of the rules is that of experiments conducted at a constant rate of elongation, but obvious modifications will occur to those interested in the other types of experiments. It also seems pertinent to observe that whenever the general criterion for the existence of a simple system (the three-element model) has been established, the single dashpot has its kinetics, however peculiar, completely bared, and the necessity for employing the rules vanishes. The concept of compound models, because of its great difficulty from the experimental standpoint and its complicated and involved nature from the theoretical standpoint, is reserved for a final note.

The typical experimental material available is a complex stress-strain curve like Fig. 13, which consists of many cycles of elongation, relaxation, and contraction. In Fig. 13 the curve is plotted, for compactness, on a load-elongation plane, rather than the alternative plot on the load-time plane. The six rules are as follows.

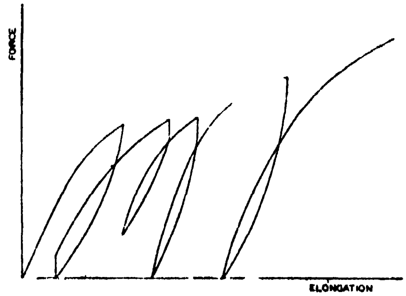


FIG. 13. A complex stress-strain curve. Relaxation is indicated by a straight vertical line.

### THE SIX RULES

I. The complex stress-strain curve is compared with simple experiments in which samples of the material under study are elongated to break at the rates of elongation under study (Fig. 14). The complex stress-strain curve at a given rate on the load-elongation plane must lie within the simple curve taken at the rate considered as an envelope (criterion I) (Fig. 15). Any portion of the complex curve lying outside of the envelope (Fig. 16) must be analyzed on the basis of a theory of thixotropy.

An obvious exception to this rule is encountered when the rate is changed rapidly enough to leave the stress-strain curve within the envelope at the old, faster rate, but outside of the envelope at the new, slower rate (Fig. 17).

II. A large number of points of no relaxation is determined. A point of no relaxation is determined by finding an elongation where the force decreases upon relaxation, then contracting the sample until a point is reached where the force increases with relaxation, eventually bracketing the force value when the initial rate of relaxation is zero (Fig. 18). Out of the totality of points so determined, a special group is defined as follows: a member is taken directly by one contraction starting from a point where the complex curve lies on its envelope. This definition will have to be approximated by doing as little searching as possible, and a certain amount of discretion must be used in placing a point in the special class.

Because there is only one defined way of achieving a member of this group from a definite elongation on the envelope, the totality of these points will form a line underneath the envelope that is a function of the elongation. Consider the difference in force between the envelope and these points. If this difference increases monotonically and becomes constant, then, in terms of spring-dashpot language, the elements that are determining the point of no relaxation are the same after this constancy has been achieved. When this condition of constancy is satisfied early in the experiment, criterion II is obeyed. The group of elements that determine the points of no relaxation can then be investigated as an entity. Before criterion II is satisfied the region, if it is satisfied ultimately, is to be treated under the rules given below concerning the return (contraction) curves. Data in this region, for most practical cases, are fragmentary and unimportant.

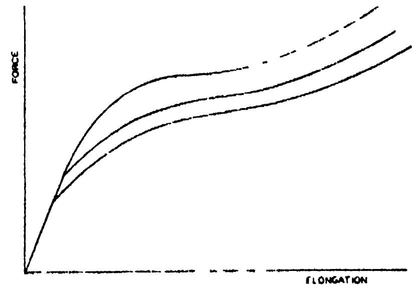


FIG. 14. Samples elongated to break at several rates.

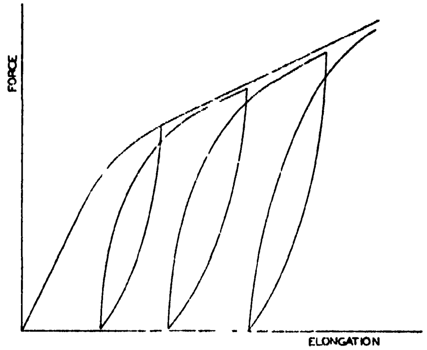


FIG. 15. A complex curve lying completely within the envelope (conforming to criterion I).

When this condition of monotonic increase and constancy is not satisfied, then some change in the determining flow process with elongation is implied.

III. *The residue of points* of no relaxation, after the special group is considered, is divided into sections totally separated from one another by any relaxation occupying a longer time than is the criterion of direction of movement, for determining the point of no relaxation. It is, of course, also separated by any determination of a point of the special class or any return to the envelope. The process described in this rule is repeated for each group of points isolated by these occurrences. Within such a section, a spring line is drawn from the terminal point of such a long relaxation or from the point of a determination of the special group, with a slope that makes it pass as closely through the points in the section (Fig. 19) as possible. If it is possible to pass this line through all the points in the section, and if it is possible to do this with the same slope in every section, then the concept of a single relaxing dashpot for determining the point of no relaxation is applicable. (Criterion III.) The example of a simple "distribution" given above is of this kind.

IV. If the previous criteria are observed, it remains to consider the shape of the return loops. A simplifying criterion is the existence of a well-defined slope in the region of no relaxation. If this constant elastic slope exists throughout the experiment, then it is established that the same totality of Hookean springs is operating throughout the experiment. (Criterion IV.) When this criterion is not observed, some condition of thixotropic flow or non-Hookean elasticity is implied.

V. When the elastic slope of rule IV is established, the working curves are transcribed, using the slope of rule III as spring line and the points of no relaxation as origin. If a well-defined working curve exists, the appropriate theory can be applied (Criterion V)

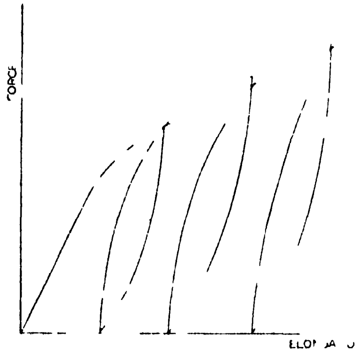


FIG. 16. A complex curve with a portion outside the envelope (violating criterion I).

VI. If the working curve is the same for elongation and contraction and the constants determined from relaxation agree with those from shape and rate change, then the simple hyperbolic-sine model can be employed. (Criterion VI.)

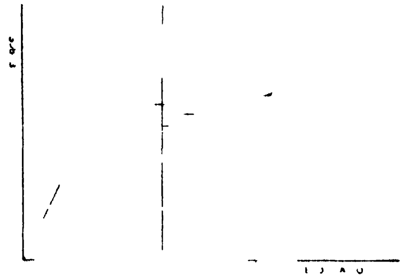


FIG. 17. The case of a change from a fast to a slow rate in which the curve can be said to obey criterion I because it approaches the envelope, although from the outside.

### NOTES AND EXCEPTIONS TO THE RULES

I. When only a certain small portion of the curve appears outside of the envelope, this part can often be neglected, and points of no relaxation taken from it excluded from the special class. Usually this "hump" is observed just as the curve reaches the envelope after a series of elongations and contractions confined to the interior of the permitted area. By a careful study of when the humps appear, a proper thixotropic scheme can be devised (see below). Although the special class of points of no relaxation is not affected, the behavior of the other points must be carefully studied in the light of the thixotropic scheme. When, however, the position of the curve is a complete departure from the envelope ("strain hardening"), an extremely thorough study of the effect of order of succession in the performing of experiments becomes necessary. This complication of experiment is required because the great simplification of the possibility of proceeding to a reproducible state (the envelope) at will, is no longer available. When the general thixotropic scheme has been devised for the envelope, it is still a much greater task to go to the return loops and points of no relaxation. There is no example where this has been done.

Certain deviations from the envelope in the sense that the complex curve lies below the envelope when it is obviously parallel to it and therefore unable to reach it, can be ascribed to true permanent set, or quasi-permanent set from a later relaxation process. Because of the usual highly non-Newtonian nature of the yielding process, it is not an important occurrence. Even when the first criterion is not satisfied, as long as the curve over a given range conforms within appropriate limits, the procedures of the following rules can be applied with caution. When the general thixotropic scheme has been elucidated,

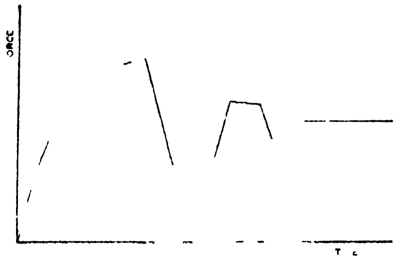


FIG. 18 The determination of a point of no relaxation.

then more general modifications of subsequent rules can be applied.

When thixotropy is established, it is quite apparent that for the envelope curves themselves whenever a dashpot is changing its rate of flow, thixotropy is likely to enter. Therefore, there might be some advantage in constructing a hypothetical, experimentally unrealizable envelope curve calculated on the basis of all of the dashpots continuously having their ultimate viscosity. The obvious reason for rejecting this procedure is the uncertainty of any theory that might be used to calculate such a curve.

The essential nature of the envelope criterion is that all curves should approach a common curve when the sample is pulled for a length of time with no reversals or halts. It is conceivable that the nature of the multiple state flow might be rheopectic, and at first the viscosity be lower than its ultimate value. No striking test for this situation is apparent, and a careful study of the consequent varying rate of approach to the envelope is required. A curve that almost parallels the envelope at a distance for a time and then ultimately approaches, indicates rheopexy, but this criterion is necessarily more vague than the corresponding one for thixotropy. When a large number of points of no relaxation are not experimentally available, then a study of the envelope curves themselves is of some use. At first, the curves at the varying rates are coincident, corresponding to the action of spring forces alone, and then they diverge as the rate-dependent dashpots come into action. The lines will then roughly parallel one another, until further yielding processes are discovered, when they will separate farther or perhaps, because of a non-Hookean parallel spring, come closer together.

II. It is to be observed that the selection of slope in rule III has no effect on the special group of points of no relaxation. The adherence to the criterion merely implies that whatever the elements that are important in the inside area, they remain the same. No assumptions of independent action of relaxation or simplicity of moving dashpot enter. When the criterion holds, it is possible to treat the single contraction curve starting at the envelope as an individual. If it, transcribed as a working curve, does not maintain constant shape while the criterion continues to hold, then a change in symmetry of potential barrier with elongation is implied.

When the compression force on the other side of the line of points in the special class is considered, it becomes quite conceivable that, while the contraction curve from the envelope to the line of the special class may be of unchanging shape and length, the prolongation of the curve may deviate from this constancy on the compressive side of this line. When this situation is encountered, it is proper to define some minor special classes, supplementing what can be called the major special class, of rule II. A suitable member would be defined as the points determined directly after going to a given line parallel to the envelope, and below the major class line. This new line would be parallel to the major line, because, by hypothesis, criterion II is already satisfied. Then if the new line of points of no relaxation is also parallel, the determining elements are identical with those determining the major line. It is apparent that very peculiar elements indeed would be implied if these lines were not parallel.

Other improper special classes can be designated by using the line of no forces as a reference. These special classes are improper because they involve both the envelope and the line of no force as references. These lines diverge as elongation increases, and so the points determined are not comparable at differing elongations. The significance of any special class is that the points it comprises are obtained from a fixed reference (the envelope) by a defined procedure. Then elongation becomes the only variable. For example, the improper class of points taken directly after a contraction from the envelope to zero force will ultimately parallel, not the envelope, but the line of zero force when the determining dashpot system is flowing at a constant rate. It is apparent that for most purposes, the major special group is sufficient and that for any situation this group is of paramount importance.

The point of no relaxation with the greatest value of force at a given elongation is attained by going into relaxation at the envelope for as short a time as is found to be necessary, contracting a very little, and looking for reversed relaxation. A similar process at zero forces finds the point with the lowest value of force.

When the second criterion fails to hold, and the envelope is a straight line when it is failing to hold, it is reasonable to assume that the cause of the failure is not the discovery of any new dashpots, but a change in the dashpots already concerned. The dashpot is of the quasi-viscous type, and the proportion of the energy dissipated to the energy stored is changing as a function of elongation, or perhaps even of history (thixotropy). Dashpot systems of this type are of fundamental importance when recycling experiments reveal dissipation with perfect elasticity (Nylon rubber).

III. While it would appear on first glance that the slope determined in III enters into the position of the points in the special group, it is nevertheless true that one of the residual points is necessary to fix this slope. The special class depends only on the envelope slope. The slope of III can often be guessed by passing the spring line through the points where the elongation and contraction curves have equal slopes (the "well-defined" slope of IV). This process depends, of course, on the adherence to criterion IV.

IV. In the case of wool, a lowering of slope because of continued working in the region of no relaxation shows that the molecular chains are resting in temporary positions of high energy when it becomes easy to coil and uncoil them. A relaxation of sufficient time restores the original slope. The initial slope is often steeper than subsequent ones. This can be ascribed to a deep-seated thixotropy, or to a non-Hookean behavior.

V. If there is definite evidence for the presence of a distribution, the transcription of a working curve will probably fail, unless the singleness of the determining dashpot is established. Then, over moderate ranges, a working curve for this single element can be established. In a recycling within a short time after being on the envelope, the yield to the envelope will be very sharp, no matter how broad the distribution, because all of the elements are in phase. Under these conditions the securing of a working curve for the determining element is facilitated. However, after a long relaxation, the yield is very gradual because the more non-Newtonian dashpots have held the compression of the associated springs. Here the hypothesis of singleness of determining dashpot is most likely to break down.

To treat the relaxation curve quantitatively, the first portion of it is ascribed to the determining element. From this portion, the constants of the determining element are estimated, and the result compared with those from shape. Then the change in envelope with rate is investigated, and the amount ascribable to the finding element subtracted. The remainder of the relaxation curve is split into enough parts to absorb the residue. This process can be made to yield a number which characterizes the distribution. The number of independently acting dashpots to which the relaxation curve must be attributed to make the non-Newtonian flow ratio, determined from speed change, come out right, is a measure of the extent of the distribution. Then, this distribution must be harmonized with the shape factor by postulating a distribution in activation energy instead of yield value. Because the determining dashpot changes with rate, this briefly indicated process must be a matter of cut and try, with a gradual assimilation of all of the data.

## NOTE CONCERNING THE TYPE OF EXPERIMENT TO BE PERFORMED

Except for certain path-finding experiments, the effort should be made to test the criteria in

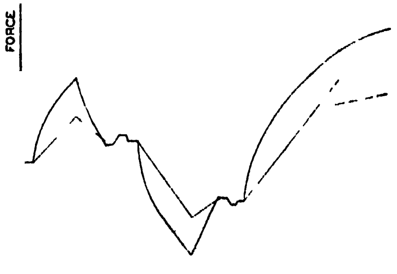


FIG. 19. The determination of the spring line.

the order given. First, a number of elongations to break, at differing rates, should be performed. These serve as general guides to the shape of the curves, and indicate the presence of the various yielding mechanisms. To find the difference in yield force because of rate, the effect of sample variation must be obviated. This is done by changing the rate, while at a point of no relaxation, on a single sample. Then the complete elongations at one rate are used to extrapolate to various elongations not covered by the complex rate experiment. The order in which the rates are applied must be permuted to determine if any strain hardening exists. (See the note to I.) If it does, then this portion of the experimental investigation must be prolonged until the rules governing the hardening can be ascertained and incorporated on a thixotropic scheme.

Then, experiments must be performed to get representative members of the special group at no relaxation points. These can be performed at one rate for each experiment, and keyed to one another by the earlier experiments at differing rates. In these experiments, the first proofs of thixotropy may be discovered, if thixotropy does not depend on backward relaxation, and manifests itself by a passing outside of the envelope. Thixotropy can be neglected until criterion II can be proved or disproved. If this criterion is obeyed, then the treatment is simplified.

Next, the experiments consisting of return cycles and the determination of the residue of relaxation points must be performed. If thixotropy is indicated, these experiments must contain many attempted returns to the envelope to elucidate the thixotropic scheme. In the presence of thixotropy, little more can be accomplished because the thixotropic working curves remain to be calculated although the principles governing the calculation are clear. However, even in the presence of thixotropy, the finding dashpot may not be thixotropic, and thus can be studied more easily if return loops starting at a thixotropic peak are not present. Last, a study of relatively long time relaxation should be undertaken.

## NOTE CONCERNING THE REPRODUCIBILITY OF THE EXPERIMENTS

Extremely poor reproducibility enormously complicates single fiber work. A trivial type is lack of scale reproducibility. If two constant-rate-of-elongation experiments differ only by the scale of the force, then a suitable area correction can be applied.

The most significant type of deviation is encountered in the early part of most curves. The first yield value may be abnormally low. This can be ascribed to greater than usual initial tension or lack of over tension ever occurring in the preparation. On the other hand, much more commonly the initial yield force may be very high, showing that at some time in the preparation the sample received an over tension that more or less permanently moved some highly non-Newtonian elements, leaving them in compression in the unstressed state. Because of the long-term relaxation, or probably more precisely creep recovery, the yield in this case is high but gradual. Because of the almost endless possibility of the concealment of progressively less Newtonian elements, it is wise to neglect and adjust for the first cycle in the early experiments. Thus, after the first good approach to the envelope, the curves will reproduce. Later on, long-time creep and relaxation studies will be required to elucidate the first part of the curve.

For this reason in analyzing stress-strain curves the position of the curve as a function of rate is known very accurately, while the position of the true spring line must be estimated by an unsure speculation. The formula for calculating $\beta$ from two rates and the position of the spring line is thereby limited in accuracy to the precision with which the spring line can be placed. It seems a logical step to replace this formula involving two rates and the spring line by the formula based on three rates alone. If $\beta$ is the value of this constant referred to the lowest rate, then for rates $m$ and $n$ times as fast, the corresponding values are $m\beta$ and $n\beta$. Thus $f_1=(1/\alpha)\,\text{arc sinh}\beta$, $f_m=(1/\alpha)\,\text{arc sinh}m\beta$, $f_n=(1/\alpha)\,\text{arc sinh}n\beta$.

The measured ratio $R_{mn}$ is defined

$$R_{mn}=\frac{f_n-f_1}{f_m-f_1}$$

$$=\frac{[\text{arc sinh}n\beta-\text{arc sinh}\beta]}{[\text{arc sinh}m\beta-\text{arc sinh}\beta]}.$$

To the approximation that

$$\text{arc sinh}\beta = \ln[\beta + (1+\beta^2)^{\frac{1}{2}}] \approx \ln 2\beta,$$

this equation becomes, for large $\beta$

$$R_{mn} = \ln n / \ln m.$$

That is, to a good approximation, this measured ratio remains independent of $\beta$ and thus this method of calculating $\beta$ fails. Hence, it is apparent that if only the experimental values of the curve envelopes are being considered, the carrying of the constant $\beta$ is redundant. The one-constant non-Newtonian element is immediately suggested. Taking the dimensional datum $f_n - f_1$ as the measured quantity, a dimensioned $\alpha$ can be directly computed from:

$$f_n - f_1 = (1/\alpha)(\text{arc sinh} n\beta - \text{arc sinh}\beta).$$

If the spring line is sufficiently far away to make $\beta$ large, then the approximation for arc sinh can be employed, and

$$f_n - f_1 = 1/\alpha_\infty \ln n$$

or

$$\alpha_\infty = (\ln n)/(f_n - f_1).$$

The symbol $\alpha_\infty$ then denotes a limiting value of dimensional $\alpha$ as $\beta$ becomes very large. It should not be confusing that $\alpha$ in reduced units reaches no limiting value but goes on increasing. The new $\alpha_\infty$ has been computed directly in the units of force used for $f_n - f_1$. The utility of this reduction in the number of constants is increased when it is applied to the more complex models for viscous elements described under the heading "The Thixotropic Viscous Element."

Any other lack of reproducibility at present is ascribed to essential differences in the samples.

Having eliminated the spring line, it is now possible to apply the formula for $\alpha$ to cases where it is known that the three-element model is inapplicable. Whenever the envelope curves taken at the varying rates are parallel, the quantity $f_n - f_1$ in the equation for $\alpha_\infty$ can be measured. The constant so obtained applies to whatever dashpots are affected by rate. Now if criterion III is obeyed, so that a single dashpot determines the major special group, then it follows that this dashpot is markedly more affected by rate than the others, and therefore the constant $\alpha_\infty$ applies almost exclusively to it. Even when the criterion has not been established, this method of dealing with a group of envelope curves is satisfying and a good approximation.

It is in reality a first viscous datum for any material, because, while the value of the yield force is a function of history and generally increases after a few cycles to meet any stress situation, the constant $\alpha_\infty$ is always applicable to define the effect of changing rate, and therefore impact, on the material.

## NOTE ON COMPOUND MODELS

### Higher Order Experiments

If in a stress-strain experiment a point of no relaxation is found and then a relaxation curve is taken starting at this point, a definite relaxation, although at a very low rate, is encountered. At a given elongation this second-order relaxation can be either upwards or downwards depending on the force on the fiber. By a judicious interpolation of relaxations the median force where this relaxation is neither up nor down, and constants for this second-order dashpot can be found. The position of the median can be followed as a function of elongation and the second order version of rule II can be used, if some definite mode of finding the median be adopted. If the position of the median is independent of the mode of finding it, at a given elongation, then the second-order analog of the three-lement model has been established. As the rudiments of a third-order investigation, the very slow relaxation of the median itself can be investigated as a function of a second-order length parameter.

### Lower Order Experiments

Although certain of the criteria may fail for the stress-strain curve as a whole, in the region of the plane bounded by the envelope a small stress-strain curve that contains no part on the envelope can be investigated. It may be found that among the easily moved elements that come into play in this region, definite yield envelopes can be identified.

The curves inside these sub-envelopes can then be investigated exactly like the gross stress-strain curve, except that experiments must be refined. The same rules have a greater chance of success in this region of less drastic forces. If various sub-sections, separated by stretches of curve on the gross envelope, yield to the same analysis, up to and including rule V, then with *these constants of a simple hunting element* sub-curve, a gross curve with a doubly compound hunting element can be defined. In fact, a gross curve that did not yield to rule II could be satisfactorily analyzed if its sub-curves yielded to the rules. The constants so obtained for the sub-curves would show how the make-up of the determining or hunting element of the gross curve varied with elongation.

## THE THIXOTROPIC VISCOUS ELEMENT

It is possible to construct a viscous element with a dependence on history of treatment, and with a state that changes in time by postulating a double (or multiple) series of potential energy curves, one of which represents the molecules in a rest state when flow is difficult, and the others repeated promoted states revealed by flow and by thermal agitation. In this development of the simple thixotropic element, the volume of flow displacement, $\alpha$, will be considered constant.

The equations developed will be for true viscous elements, that is, the external force will be the sole agent causing flow, and when it is removed flow will stop. Several forbidden systems will be discussed to show that this is not always the case. The second condition for a true viscous element, namely, the identity of behavior at any value of the flow measuring coordinate, will be observed in all cases. When several intercommunicating positions of rest (potential wells) at one value of the flow coordinate are available, then the rate of flow of the system is not a simple function of the force on the system alone. The population of the wells becomes a factor.

To take a specific example, if two positions of rest are available, transformations can conceivably be of importance along all of the paths indicated in Fig. 20 where the circles designate positions of rest and the lines, transformations. Transformations along these lines are governed by laws of the type

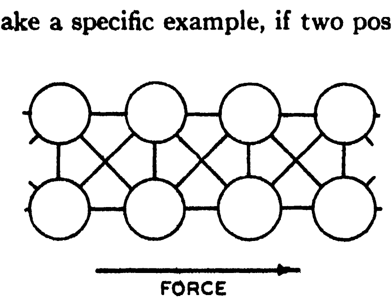


FIG. 20. A schematic representation of a complex flow system. The circles denote positions of rest while the lines represent flow processes.

$$K=(kT/h)(F^{\ddagger}/F_1)e^{-\epsilon/kT},$$

where $K$ is a "rate constant." There will be two of these equations for each connecting path. The total rate of shear is the sum of the rates of reaction in the forward direction minus the sum of the rates in the backward direction. If $n$ is the population of the top state, and $1-n$ the population of the bottom state, then similarly $dn/dt$ equals the sum of the $k$'s going upward minus the sum of the $k$'s going down.

The equations for the complex two-position model could be written out, but they quite obviously contain too many parameters to be readily used. Two principles hold: First, the assignment of constants must be symmetrical with respect to plus and minus force, or otherwise, when the force is removed from the element and the populations are in any state but the equilibrium one, there will be residual motion associated with the assumption of the equilibrium state. Second, if this symmetrical assignment is made, the force will exert no influence on the populations (the populations will remain at rest equilibrium) unless the potential barriers of the diagonal transformations are unsymmetrical.

In order to simplify the treatment, certain transformations are given sufficiently high activation energies to make them unimportant. This is symbolized in the diagram by the absence of the line representing this transformation. This transformation then is "forbidden" in both directions.

It is contrary to the principle of detailed balancing to forbid a transformation in one direction alone. A satisfactory substitute is found in the completely unsymmetrical barrier. In the normal symmetric barrier if the forward rate is increased by a multiplicative factor $\Phi\equiv e^{\phi}\equiv e^{\alpha f}$, then the backward reaction is retarded by the same factor. In the case of the completely unsymmetrical barrier, the reaction in one direction is affected by $\Phi^2$, while the reaction in the other direction is unaffected. This is diagrammatically expressed by a line with an arrow head on the end toward which reaction is affected by the force.

The simplest true viscous element will now be formulated (Fig. 21). The undisturbed rate along the horizontal arrow will be denoted by $K$. Then

$$\begin{aligned} k_{\longrightarrow} &= (1-n)K\Phi, \\ k_{\longleftarrow} &= (1-n)K\Phi^{-1}, \\ k\nearrow &= (1-n)KD, \\ k\nwarrow &= (1-n)KD, \\ k\searrow &= nKD\Phi^2, \\ k\swarrow &= nKD\Phi^{-2}, \end{aligned} \tag{49}$$

when $D$ is a factor for the difference between the diagonal passage rate and the horizontal rate. Then the rate of elongation is

$$dl/dt = K[(1-n)(\Phi-\Phi^{-1})+nD(\Phi^2-\Phi^{-2})], \tag{50}$$

and

$$dn/dt = KD[2(1-n)-n(\Phi^2+\Phi^{-2})]. \tag{51}$$

In these equations it has been assumed that the two available positions are at equal energies, and thus at equilibrium with no force applied, $n=0.5$.

To calculate working curves, the condition of experiment must be specified. At a constant rate of elongation from (2)

$$\begin{aligned} dl/dt &= \rho-(1/k_1)(df/dt) \\ &= K[(1-n)\sinh\alpha f+nD\sinh 2\alpha f)], \end{aligned} \tag{52}$$

where the factor $\frac{1}{2}$ in the definition of sinh has been adsorbed in $K$, and

$$dn/dt = KD[(1-n)-n\cosh 2\alpha f]. \tag{53}$$

In reduced units

$$\beta - d\phi/d\tau = (1-n)\sinh\phi + nD\sinh 2\phi. \tag{54}$$

Now, if the constants are known, a point-by-point calculation will give the working curve. To calculate the constants from the experimental working curve is more difficult.

If experiments are available at high rates of shear, when the ultimate population of the upper state $n_\infty \approx 0$, then it is simple to calculate all of the constants but $D$ from the law governing the change of yield force with rate of experiment, utilizing the ultimate force values at two or more rates of experiment. At high rates it can be assumed that when the maximum in force occurs, sufficient reaction has not taken place to alter $n$ appreciably from its initial value if $D$ is small. If this value is the equilibrium value, then at the point of maximum force $n=\frac{1}{2}$. If $D$ is very small, then

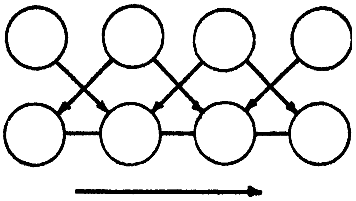


FIG. 21. The simplest true viscous element with two positions of rest.

$$\beta = \tfrac{1}{2}\sinh\phi_{\max}. \tag{55}$$

Now, by hypothesis

$$\beta = \sinh\phi_\infty. \tag{56}$$

Because $\beta$ is large, by hypothesis

$$\begin{aligned} \phi_{\max} &= \ln 4\beta, \\ \phi_\infty &= \ln 2\beta, \end{aligned}$$

and if

$$r = \phi_{\max}/\phi_\infty$$

then

$$r = \ln 4\beta/\ln 2\beta = 1+(\ln 2/\ln 2\beta). \tag{57}$$

If the contribution of $D$ is of importance

$$\beta = \tfrac{1}{2}\sinh\phi_{\max} + 2D\sinh 2\phi_{\max}, \tag{58}$$

and replacing hyperbolic sines with exponentials and using

$$\begin{aligned} \phi_{\max} = r\phi_\infty &= \ln(2\beta)^r, \\ 4\beta &= (2\beta)^r + D(2\beta)^{2r}, \end{aligned}$$

or

$$D = 4\beta - (2\beta)^r/(2\beta)^{2r}. \tag{59}$$

It is quite apparent that a model wherein the lowest state $n=0.5$ is not adequate except when thixotropy is slight. The largest permissible value of $r$ is 2, and practical values will be much lower because $\beta$ is usually high.

If it is desired to express the results solely in terms of force difference, analagous to the one constant treatment of non-Newtonian viscosity above, the argument is as follows:

$$\begin{aligned} \beta &= \sinh\phi_\infty, \\ \beta &= \tfrac{1}{2}\sinh\phi_{\max} + \tfrac{1}{2}D\sinh 2\phi_{\max}. \end{aligned} \tag{60}$$

Replacing the hyperbolic sine with the exponential form

$$\begin{aligned} 2\beta &= e^{\phi_\infty}, \\ 4\beta &= e^{\phi_{\max}} + De^{2\phi_{\max}}. \end{aligned} \tag{61}$$

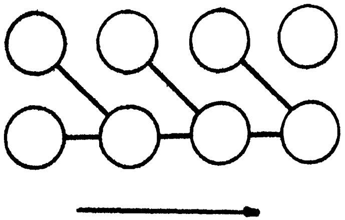


FIG. 22. A viscous system of the forbidden type

Then

$$\phi_\infty = \ln 2\beta,$$

$$\phi_{\max} = \ln\left[\frac{(1+16D\beta)^{\frac{1}{2}}-1}{2D}\right],\tag{62}$$

$$\Delta f = \frac{1}{\alpha}(\phi_{\max}-\phi_\infty) = \frac{1}{\alpha}\ln\left[\frac{(1+16D\beta)^{\frac{1}{2}}-1}{4D\beta}\right];$$

whence

$$D\beta = [2-e^{\alpha\Delta f}]/[e^{2\alpha\Delta f}].\tag{63}$$

This depends on rate, and the final expression of the answer takes the form

$$D/K = \rho[2-e^{\alpha\Delta f}]/[e^{2\alpha\Delta f}]\tag{64}$$

when $\alpha$ comes from the change in $f_\infty$ with rate, and $\rho$ is the rate of elongation. This formula applies only when $e^{\alpha\Delta f}<2$ as otherwise the model is invalidated.

This limitation could be removed by placing the upper state at an advantage thermodynamically, but this modification would introduce another constant. A permissible system, having no more constants, which does not have $n_0=0.5$ can be designed by modifying the previous system by making $D$ a difference in well energy rather than an activation energy factor. The activation energy for the diagonal process, from the bottom state, will be the same as that for horizontal flow along the bottom states. The equations expressing this are

$$\begin{aligned}
k_{\longrightarrow} &= (1-n)K\Phi,\\
k_{\longleftarrow} &= (1-n)K\Phi^{-1},\\
k\nearrow &= (1-n)K,\\
k\nwarrow &= (1-n)K,\\
k\searrow &= nKD\Phi^2,\\
k\swarrow &= nKD\Phi^{-2},\\
1-n_0/n_0 &= D,\\
n_0 &= 1/D+1.
\end{aligned}\tag{65}$$

The equation for $dl/dt$ remains the same. But repeating the approximation for

$$\begin{aligned}\beta &= (1-n_0)\sinh\phi_{\max}+n_0D\sinh 2\phi_{\max}\\ &= (D/D+1)(\sinh\phi_{\max}+\sinh 2\phi_{\max}).\end{aligned}\tag{66}$$

Then

$$D = [(2\beta)^{2r-1}+(2\beta)^{r-1}-1]^{-1}.$$

In terms of force differences

$$2\beta = e^{\phi_\infty},$$

$$2\beta = (D/D+1)(e^{2\phi_{\max}}+e^{\phi_{\max}}),\tag{67}$$

whence

$$\phi_\infty = \ln 2\beta,$$

$$\phi_{\max} = \ln\frac{[1+8(D+1/D)]^{\frac{1}{2}}-1}{2},\tag{68}$$

$$\alpha\Delta f = \tfrac{1}{2}\ln 2/\beta D,\quad (\beta \text{ is very large}),$$

$$D/K = 2\rho e^{2\alpha\Delta f}.\tag{69}$$

A system which is of interest is shown in Fig. 22. It is of the forbidden unsymmetrical type and, for this reason, can function without unsymmetrical potential barriers. The equations governing its motion are

$$\begin{aligned}
k_{\longrightarrow} &= (1-n)K\Phi,\\
k_{\longleftarrow} &= (1-n)K\Phi^{-1},\\
k\searrow &= nKD\Phi,\\
k\nwarrow &= (1-n)KD\Phi^{-1},\\
dl/dt &= K[(1-n)(\Phi-\Phi^{-1})\\
&\qquad +nD\Phi-(1-n)D\Phi^{-1}],\\
dn/dt &= KD[(1-n)\Phi^{-1}-n\Phi]
\end{aligned}\tag{70}$$

At a constant rate of elongation, when $d\phi/dt=0$,

$$\begin{aligned}\beta &= (1-n)\sinh\phi_{\max}+nD(e^{\phi_{\max}}/2)\\ &\qquad -(1-n)D(e^{-\phi_{\max}}/2).\end{aligned}\tag{71}$$

If $n$ has its equilibrium value $\frac{1}{2}$, following a similar procedure gives

$$\beta = \tfrac{1}{2}\sinh\phi_{\max}+\tfrac{1}{2}D\sinh\phi_{\max},$$

or

$$\begin{aligned}2\beta &= (1+D)\sinh\phi_{\max} = (1+D)\beta^r,\\ D &= (2\beta/\beta^r)-1.\end{aligned}\tag{72}$$

If $n$ has the extreme value, unity,

$$\beta = \frac{De^{\phi_{\max}}}{2} = \frac{D(2\beta)^{r_{\max}}}{2},$$

$$D = (2\beta)^{1-r_{\max}}.\tag{73}$$

For this third "forbidden" case the argument based on force differences is as follows:

$$\begin{aligned} 2\beta &= e^{\phi_\infty}, \\ 2\beta &= \tfrac{1}{2}e^{\phi_{\max}} + \tfrac{1}{2}De^{\phi_{\max}}, \qquad \text{if } n=\tfrac{1}{2}, \\ \phi_\infty &= \ln 2\beta, \\ \phi_{\max} &= \ln(4\beta/1+D), \\ -\alpha\Delta f &= \ln(1+D/2), \\ D &= 2e^{-\alpha\Delta f}-1. \end{aligned}$$

If $n=1$,

$$\begin{aligned} 2\beta &= De^{\phi_{\max}}, \\ \phi_{\max} &= \ln 2\beta/D, \\ -\alpha\Delta f &= \ln D, \\ D &= e^{-\alpha\Delta f}. \end{aligned} \tag{74}$$

The simplest of all models is an improper one, containing an unsymmetrical barrier (Fig. 23) The equations governing this model are, assuming all activation energies and energy levels equal,

$$\begin{aligned} k_{\rightarrow} &= (1-n)K\Phi, \\ k_{\leftarrow} &= (1-n)K\Phi^{-1}, \\ k_{\searrow} &= nK, \\ k_{\nwarrow} &= (1-n)K\Phi^{-2}, \\ dl/dt &= K[(1-n)(\Phi-\Phi^{-1})+n-(1-n)\Phi^{-2}]. \end{aligned} \tag{75}$$

At a constant rate of elongation, when $\phi$ is maximized

$$\beta = (1-n)\sinh\phi_{\max} + n - (1-n)e^{-2\phi_{\max}}. \tag{76}$$

Because there is no extra constant, $n$ can be fixed as a function of $r$

$$n = 1 - 1/(2\beta)^{r-1}. \tag{77}$$

For the last and simplest case, the following argument holds in terms of force differences:

$$\begin{aligned} 2\beta &= (1-n)e^{\phi_{\max}}+n, \\ \phi_{\max} &= \ln 2\beta/1-n, \\ -\alpha\Delta f &= \ln(1-n), \\ n &= 1-e^{-\alpha\Delta f}. \end{aligned} \tag{78}$$

These equations will be employed in the analysis of some typical thixotropic curves.

## Rules for Analyzing the Working Curves

The examples of special two-position working curves given above make clear the principles governing the calculation of curves of this type, and indicate methods of calculating the constants when the curve is found experimentally.

The complexity introduced with only one new constant, and in one case with no new constants, is apparent. Realizing that these cases were selected for their simplicity rather than for their reality, it becomes apparent that the selection of the particular multiple-position model is a difficult task.

A series of rules is set up for analysis, assuming that the over-all analysis has been carried to such a point that the working curve of the element in question is known at least to an extent of uncertainty no greater than the presence of an unknown additive factor.

I. The working curve must have a constant slope at and near the origin (criterion I).

II. The working curve is investigated at several velocities and the behavior of the ultimate value of the force investigated. If this force is a function of the velocity alone, strain-hardening is absent (criterion II).

Using these ultimate forces, a value of $\beta$ is computed for the bottom process of flow. Similarly, the behavior of the hump itself is investigated as a function of the treatment after leaving a steady value of ultimate force. The shape should depend only on this recent history (criterion IIA).

III. The difference between the maximum force (at the hump) and the ultimate force is investigated after cycles of elongation and contraction followed by periods of no force. If the largest maximum force always follows the longest period of no force, other conditions being equal, then a symmetrical model is indicated (criterion III).

IV. The maximum value of $r$ is compared with the value $r=\ln 4\beta/\ln 2\beta$. If $r$ is ever greater than this value, it is necessary to employ a model with the difficult-to-flow state thermodynamically favored.

V. If $r$ is not greater than $\ln 4\beta/\ln 2\beta$, the

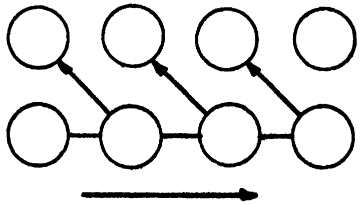


FIG. 23. The simplest (improper) model.

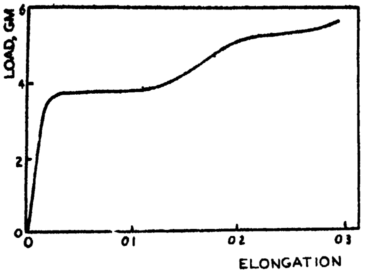


FIG. 24. Elongation *vs.* load to break for the chicken keratin fiber, diameter of 90$\mu$, $\rho=0.402\times10^{-3}$

single constant can go into an activation energy factor to hinder the diagonal reaction.

These rules have been carried only far enough to provide for a single new constant. They could be expanded to include tests of a quantitative nature to employ more constants. No radically different behavior would be encountered within the framework of two equilibrium positions.

If the first criterion is disobeyed; it becomes necessary to employ a "thixotropic spring" model; if the second criterion is disobeyed, strain hardening is involved. If the third criterion is disobeyed, an unsymmetrical model is implied, and, in any one of the three cases, the simple definition of a viscous element fails.

## Experiments with Chicken Keratin

When Dr. Harold P. Lundgren of the Department of Agriculture's Western Regional Research Laboratory visited the Textile Research Institute Laboratories, he left a sample of the fiber prepared from chicken feather keratin in his laboratory. This material shows many interesting properties, including flow thixotropy, constant force springs, and the presence of a complex distribution of relaxation times. The fiber is very sensitive to humidity. When it is wet, it is very soft. At 40 percent relative humidity, it has a stress-strain curve that does not have any thixotropic hump but which shows very clearly the constant force spring. Figure 24 shows an elongation to break. At 50 percent humidity the thixotropic hump appears, and all of the curves of this section except Fig. 24 were made at this humidity. The region of flatness is moved up and is correspondingly longer than the flat region of Fig. 24. At 60 percent relative humidity, the fiber becomes so stiff that it often breaks before the yield point; therefore, long-range experiments become impossible except in the region below the yield point.

Because the fiber was prepared in a small-

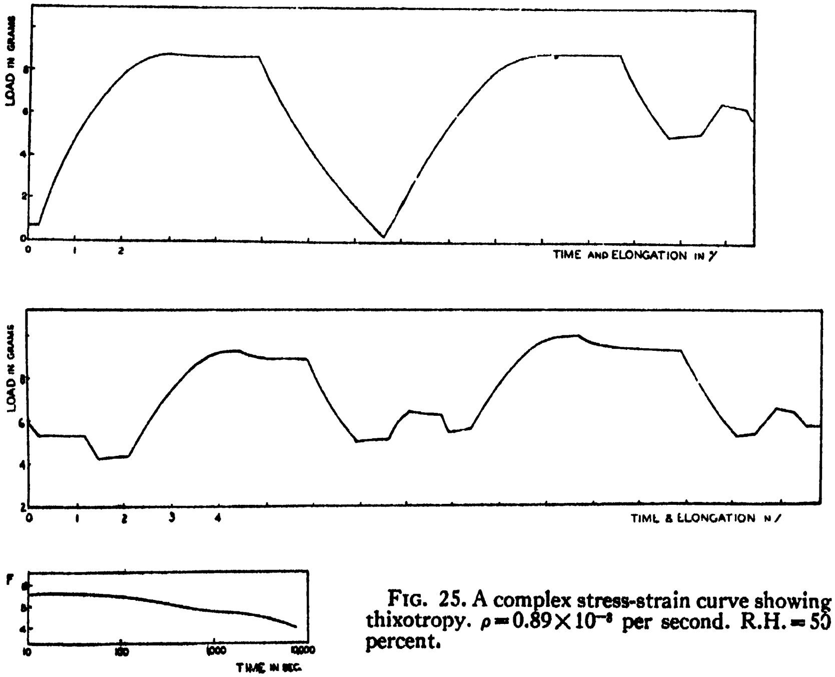


FIG. 25. A complex stress-strain curve showing thixotropy. $\rho=0.89\times10^{-3}$ per second. R.H.=50 percent.

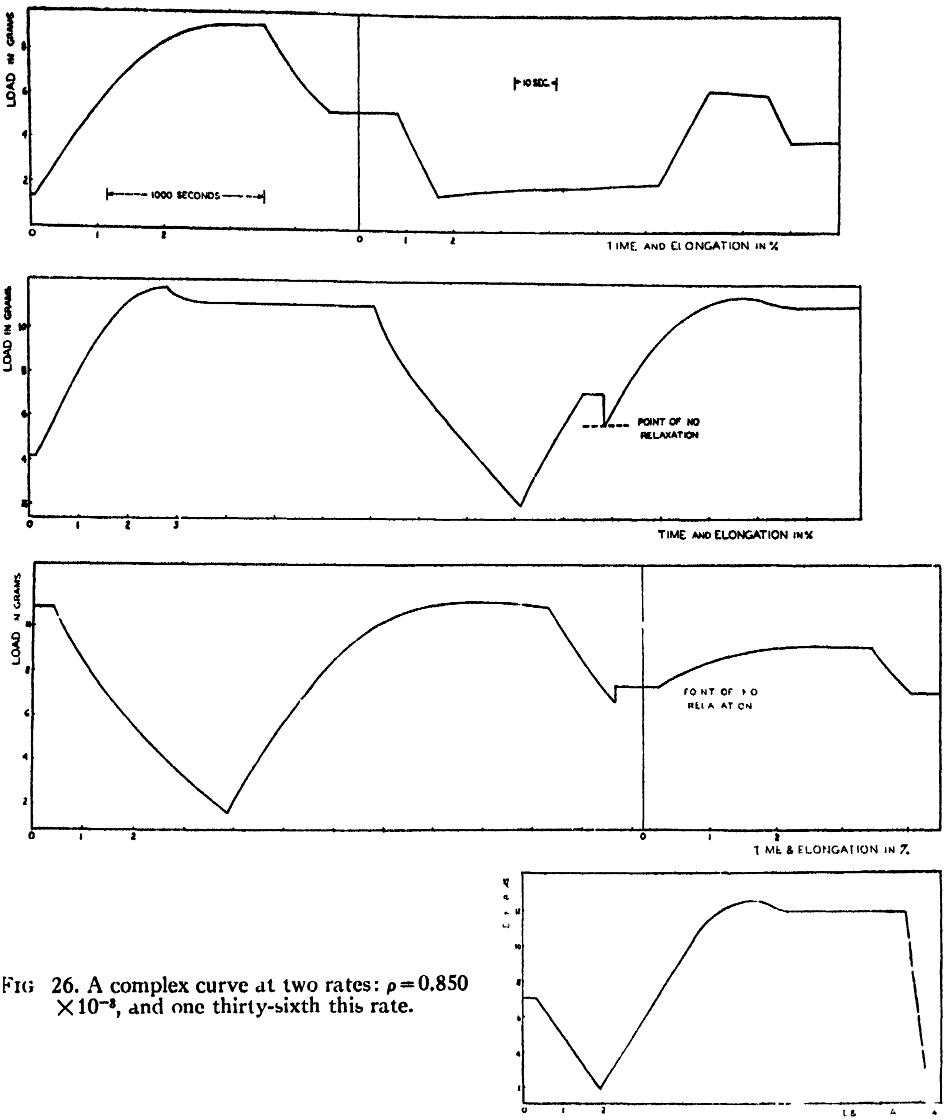


FIG. 26. A complex curve at two rates: $\rho=0.850 \times 10^{-3}$, and one thirty-sixth this rate.

scale experimental fashion, it was quite irregular in behavior. Therefore an investigation to determine the exact components of its visco-elastic system would have been very trying. For that reason, the method of expressing constants that depends only on differences in force and not on the position of the ultimate spring line will be employed.

Figure 25 shows results of an experiment. The first panel shows that in the beginning of the experiment there was very little hump in the curves. After working some time to determine a point of no relaxation, a pronounced hump of about one gram has developed. A situation of this nature shows that the unsymmetrical "forbidden" type of element must be employed. As will be shown later, the meaning of the unsymmetrical element can be expressed in configura-

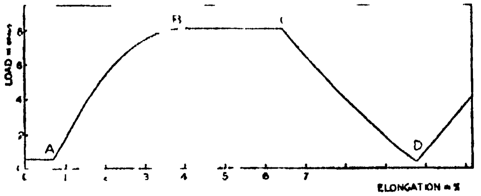


FIG. 27. A portion of the stress-strain curve for chicken feather keratin. The diameter of the fiber is approximately $90\mu$, and $\rho=0.835\times 10^{-3}$ per second, R.H. = 50 percent.

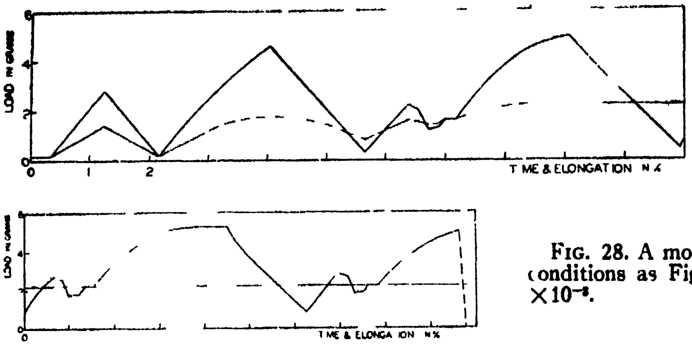


FIG. 28. A more complex curve prepared under similar conditions as Fig 27 except R.H.=60 percent, $\rho = 0.40 \times 10^{-6}$.

tion space. Hence, there is no uncertainty about the use of such an element.

Now a very long experiment with many differing treatments would be necessary to justify a very complex picture of the flow system in configuration space. Because of the fragility of the fiber, this is very difficult to realize. The two-constant improper thixotropy theory is adequate to explain the curves here which are of a fair degree of complexity. The last panel of Fig. 25, in which the relaxation over a long period of time, starting with a point of no (instantaneous rate of) relaxation, shows clearly that a complex distribution of relaxation times enters into a complete description of the fiber. By using the one-constant method, the analysis of this feature of the curves can be avoided and attention focused on the problem of thixotropy

The stress-strain curve of Fig. 26 is complex enough to allow a computation of some constants. The change in ultimate force caused by a rate change of a factor of 36 was 2 0 grams. Thus $\alpha_\infty = \ln 36/2.0 = 1.8$ per gram. $\Delta F$ is very close to 1.0 gram when carefully taken from a point of no relaxation. Therefore, $\alpha\Delta F = 1.8$, $e^{-\alpha\Delta F} = 0.17$. Because this is a rather small quantity, denoting a large thixotropy, the over-simplification of the simplest forbidden case is not appropriate. The other forbidden case discussed in the text had an equilibrium value of $n_0$ of one-half and is similarly only slightly thixotropic.

A slight modification of this treatment can be made in which, instead of the rate of flow between states differing by a factor $D$ from the forward and backward rates, the inaccessible state is made a factor $D$ lower than the state when forward flow is direct. That is the two states differ by $\Delta F'$, and $D = e^{-\Delta F'}/RT$. Then the equilibrium value

$$n_0 = 1/D+1,$$

and in terms of the infinite force expressions one finds by arguments similar to those given in the section above

$$D/D+1 = 2e^{-\alpha\Delta F}. \qquad (79)$$

From this experiment, at 300°K, the difference in energy is found to be 1410 calories.

## CONSTANT FORCE SPRINGS

The only requirement that a true spring must satisfy is that it should store energy completely reversibly. This in turn means that the energy of a spring must be a monotonic function of the elongation. The force on the spring is the slope of the energy function, and it is expected that the force can rise, go through a maximum, and descend to zero, but never become negative. Although it is easy to construct springs of this sort artificially, there are few illustrations of them in nature.

In the analysis of the stress-strain curves of protein fibers it has been found necessary to employ non-Hookean springs in which, if it does not descend, it reaches a constant value. In Fig. 27 a portion of a stress-strain curve for the chicken feather keratin fiber is shown. At point $B$ the fiber is completely yielded to zero slope. If at point $B$ the direction of elongation had been reversed, then the contraction would have taken place quite reversibly but not to such an extent that the elongation was any less than that at point $A$. Indeed, as the strain at the reversal was decreased, recovery would asymptotically approach $A$ from the $B$ side. In the actual experiment, the reversal of direction was not made at point $B$, but the fiber was elongated at constant force, for several percent of elongation, to point *C*. Because there was no increase in force, one would expect that the entire portion of the curve from *B* to *C* would be completely dissipative and, therefore, all reversal curves from this point onward would be superposable. That this is not the case is manifest in that the elongation *CD* is greater than *BA*. That is, between *B* and *C* energy was being stored at a constant rate of elongation. This shows that a spring system was present that was operative at constant force. In Fig. 28 a more complex curve of the same type is analyzed on the basis of the three-element model incorporating for the spring in parallel a non-Hookean element of the type described. It is quite apparent that the spring cannot operate at constant force throughout its entire range, but that it must start at some Hookean slope. That slope is determined here by connecting the first point of no relaxation with the origin using a line of the proper constant slope. It is to be noted that even if at times the resulting line passes above the ultimate force of the spring, as long as the first point of no relaxation is sensibly below this ultimate, then the determined slope is valid below the ultimate, and must be rounded off when it approaches the ultimate. This is shown in Fig. 28 by the dashed lines. The working curve, which is well defined for the whole experiment, is shown in Fig. 29. However, it is known that this fiber cannot be successfully represented under every condition by the three-element model and obviously all of the points of no relaxation determined belong to the special class.[3] The three-element model is quite satisfactory over this particular range and any constants determined would be average values of some unspecified nature.

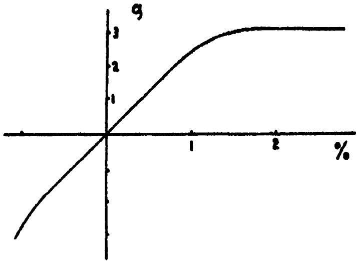


FIG. 29. The working curve for Fig. 28.

The important point, however, is that no combination of non-Newtonian, unsymmetrical and/or thixotropic elements coupled with Hookean springs could reproduce this behavior. A spring operating at *constant force* is definitely demonstrated. These springs obviously could not exist uncoupled in the free state because the elongation at the ultimate force would be undefined. But when they are coupled with viscous elements this difficulty is removed.

### ACKNOWLEDGMENTS

I wish to thank Mr. Charles H. Reichardt for valuable suggestions, and Dr. J. H. Dillon for reading the manuscript and clarifying the presentation.

# A Magnetron-Resonator System

E. C. Okress
*Westinghouse Electric Corporation, Lamp Division, Bloomfield, New Jersey*

(Received May 26, 1947)

The present exposition deals with the fundamental multiplet spectrum of a multi-sectorial cavity or vane-type magnetron and specific applications involving the evaluation of the $\pi$-mode wave-length thereof. The consequence of mutual coupling between the sectorial cavities through the interaction space of an infinitely long anode is evaluated for the linear case and subsequently applied to the cylindrical case. Although both the qualitative and quantitative results are reasonable in the limit for relatively very long anodes, in practice the latter are relatively short and since the mode separation, order thereof, and hence the component wave-lengths are more or less influenced by the actual relative anode and end cavity lengths at each end of the anode, the results naturally differ from "cold" test data. Predictions of $\pi$-mode wave-length have been obtained to about 10 percent, depending upon the degree to which the parameters of the actual vane magnetron approximate the assumptions on which the present treatment is based. With the aid of "cold" test data the preliminary theoretically derived resonator geometry may be appropriately "trimmed" to produce the exact $\pi$-mode wave-length desired.

Included in this article is a discussion of the need and function of straps of the resonator system and dimensional scaling, which preserve certain desirable properties of the prototype at the new wave-length, or operating point, in terms of magnetron performance-chart parameters. Finally, in Part III, a brief consideration is allotted to the derivation of the rectangular cavity or slot-type resonator system from the more general relation for the sectorial cavity in this article.

## I. INTRODUCTION

THE subject of this paper is based on a Westinghouse radar report[1] concerning an investigation of the symmetrical multi-sectorial-type cavity, commonly referred to as the vane magnetron illustrated essentially by Fig. 3 and practically by Fig. 9.

The analytical evaluation of the wave-length spectrum of the fundamental multiplet for the resonator system of the vane magnetron was initiated some time prior to its invention,[2] with the original object of developing a cheaper, lighter, more readily manufacturable product possessing excellent electrical properties, in contrast to the parallel slot or hole- and slot-type of resonator systems. Eventually these objectives were realized during the war, aided by the cooperative efforts of other laboratories; the vane magnetron was and is being continually introduced to many important uses to which it is especially suited.

Prior to August 1942 no results on the evaluation of the principal wave-length of the vane magnetron were known or available, so that in order to establish some basis for a preliminary design of a 10-centimeter, 8-vane magnetron the aforementioned work[1] was initiated and reported upon.

M.K.S. units are used throughout this paper unless otherwise noted.

It is appropriate to point out that only the travelling-wave magnetron oscillator is considered in this article. These oscillations are also referred to in the pre-war literature, as the transit time oscillation of higher order and type $B$. The characterization of this oscillator lies in the fact that the electronic interaction and phase relations are incorporated with the tangential and radial components of the rotating electromagnetic field in the interaction space, between the cathode and anode, respectively, in accordance with the Slater, Hartree, and Block theories of oscillation.

## II. THE RELATION FOR THE NORMAL MODES OF OSCILLATION

Consider an isolated, infinitely long sectorial cavity, with metallic confining walls, with a longitudinal slot adjoining the cylindrical hole shown in Fig. 1. Note that this cavity represents the generalization of the rectangular case to be commented on later. A mode of oscillation of future utility is chosen such that the associated components of the electromagnetic field in the sectorial cavity satisfy the conditions:

$$E_r = E_z = 0,$$

$$H_r = H_\phi = 0,$$

$$E_\phi = E_\phi(r),$$

$$H_z = H_z(r).$$

---

[1] C. C. Wang and E. C. Okress, "The vane type magnetron," Radar Report B-1, Bloomfield-29, August 1942.

[2] Patent applied for.

Since Maxwell's equations in cylindrical coordinates may be expressed as:

$$\frac{\partial E_z}{r\partial\phi}-\frac{\partial E_\phi}{\partial z}=-i\omega\mu H_r,\quad \frac{\partial H_z}{r\partial\phi}-\frac{\partial H_\phi}{\partial z}=(\rho+i\omega\epsilon)E_r,$$

$$\frac{\partial E_r}{\partial z}-\frac{\partial E_z}{\partial r}=-i\omega\mu H_\phi,\quad \frac{\partial H_r}{\partial z}-\frac{\partial H_z}{\partial r}=(\rho+i\omega\epsilon)E_\phi,$$

$$\frac{\partial}{r\partial r}(rE_\phi)-\frac{\partial E_r}{r\partial\phi}=-i\omega\mu H_z,$$

$$\frac{\partial}{r\partial r}(rH_\phi)-\frac{\partial H_r}{r\partial\phi}=(\rho+i\omega\epsilon)E_z,$$

which, for the conditions enumerated, become

$$\frac{\partial}{\partial r}(rE_\phi)=-i\omega\mu rH_z, \tag{1a}$$

$$-\frac{\partial}{\partial r}H_z=i\omega\epsilon E_\phi, \tag{1b}$$

for a perfect dielectric medium in the cavity of Fig. 1. The angular frequency, velocity of light, permeability, and dielectric constant are denoted by $\omega$, $c$, $\mu$, and $\epsilon$, respectively. By eliminating $E_\phi$ in Eqs. (1a) and (1b), Bessel's equation is obtained in the form

$$\frac{\partial^2}{\partial r^2}H_z+\frac{1}{r}\frac{\partial H_z}{\partial r}+k^2H_z=0, \tag{2a}$$

where the wave number $k=\omega(\epsilon\mu)^{\frac{1}{2}}=\omega/c$.

By a similar process, eliminating $H_z$ from Eqs. (1a) and (1b), is obtained Bessel's equation in another form

$$\frac{\partial^2 E_\phi}{\partial r^2}+\frac{1}{r}\frac{\partial E_\phi}{\partial r}+\left(k^2-\frac{1}{r^2}\right)E_\phi=0. \tag{2b}$$

In compact form the complete solution of (2a) may be written as:

$$H_z=aJ_0(kr)+bN_0(kr), \tag{3a}$$

wherein $J_0(kr)$ and $N_0(kr)$ denote the usual Bessel functions of zero order, first and second kinds, respectively.

Similarly in compact form the complete solution of (2b) or the result of the substitution of (3a) in (1b) may be written as:

$$\frac{i}{z_0}E_\phi=aJ_1(kr)+bN_1(kr), \tag{3b}$$

wherein $J_1(kr)$ and $N_1(kr)$ denote the usual Bessel functions of the first order, first and second kinds, respectively. $z_0$ denotes the intrinsic impedance of the confined space, $(\mu/\epsilon)^{\frac{1}{2}}$.

Relations (2a) and (2b) constitute all possible solutions of (1a) and (1b).

Now, applying the boundary conditions at

$$r=r_b,\quad E_\phi=0,$$
$$r=r_a',\quad H_z=0,$$

wherein $r_a'$ denotes the uncoupled sectorial cavity radius, to (3a) gives

$$H_z=aJ_0(kr_a')+bN_0(kr_a')=0, \tag{4a}$$

and to (3b) gives

$$\frac{i}{z_0}E_\phi=aJ_1(kr_b)+bN_1(kr_b)=0. \tag{4b}$$

So that from (4a) and (4b) $T_0(kr_a)$ and $T_1(kr_b)$ are defined and related at $r=r_a'$ by:

$$-\frac{J_0(kr_a')}{N_0(kr_a')}=T_0(kr_a')=\frac{b}{a}=-\frac{J_1(kr_b)}{N_1(kr_b)}=T_1(kr_b),$$

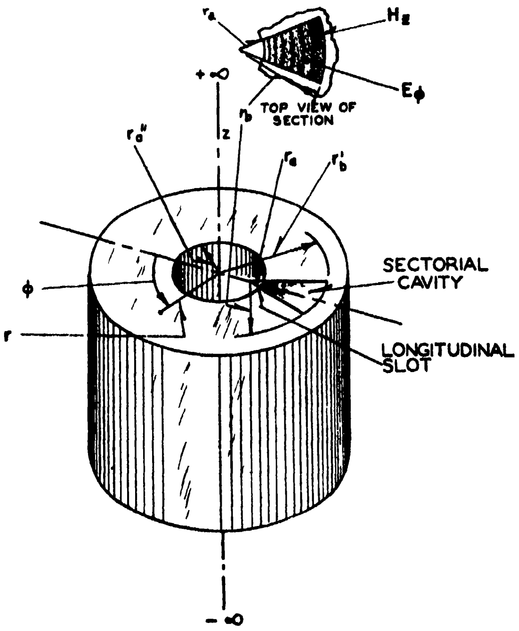


FIG. 1. A section of an infinite length sectorial cavity and adjoining chamber.

or

$$T_1(kr_b)=T_0(kr_a'). \tag{4c}$$

Equations (3a) and (3b) may be rewritten, with $E_\phi=E_{r_a}$ (and $H_z\neq 0$) at $r=r_a$, as

$$z_0H_z=E_{r_a}\cdot i\cdot\frac{J_0(kr)+T_1(kr_b)N_0(kr)}{J_1(kr_a)+T_1(kr_b)N_1(kr_a)}, \tag{5a}$$

$$E_\phi=E_{r_a}\cdot\frac{J_1(kr)+T_1(kr_b)N_1(kr)}{J_1(kr_a)+T_1(kr_b)N_1(kr_a)}, \tag{5b}$$

where $r_a$ represents the radius of the sectorial cavity coupled to the interaction space such that $(r_a-r_a')=\Delta/k$, where $\Delta$ is a small quantity.

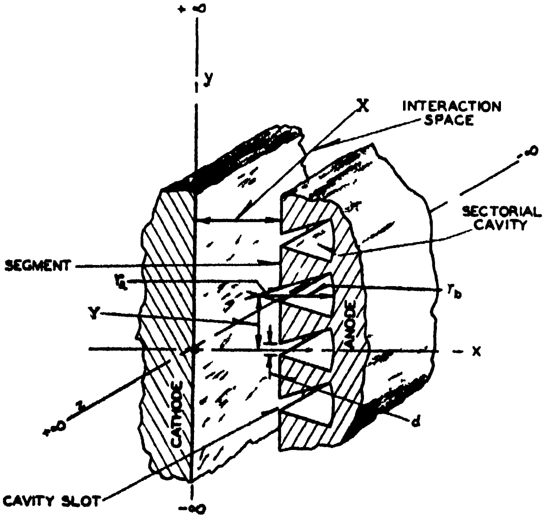


FIG. 2. A section of infinite linear multi-sectorial cavity magnetron.

Now consider the linear magnetron with an array of sectorial resonators as shown in Fig. 2. The electromagnetic fields in the interaction space (between the cathode and anode along the $y$ axis) must be constructed, and then $E_y$ and $H_z$ matched at the slots. The match or continuity of $E_y$ at the slot is exact, whereas for $H_z$ only an approximate continuity condition will be resorted to, although this is quite satisfactory for narrow slots ($d\ll Y$). In the interaction space a solution of Maxwell's equations,[3] for the electromagnetic field, of the type

$$E_y=\sum_{m=0,\pm1\cdots}-ik_{xm}\times\frac{c^2}{\omega}B_m\sinh k_{xm}Xe^{i(k_{ym}y-\omega t)}, \tag{6a}$$

$$B_z=\sum_{m=0,\pm1}B_m\cosh k_{xm}Xe^{i(k_{ym}y-\omega t)} \tag{6b}$$

is used. $k_x$ and $k_y$ denote the magnitudes of the imaginary and real components of the complex propagation vectors, $k_i$ and $k_r$, along the abscissa and ordinate in Fig. 2, respectively. The coefficient $B(=\mu H$ in rationalized M.K.S. units) denotes the magnitude of the real component of the flux density normal to both components of the complex propagation vectors. The relation between $k_{ym}$ and $k_{xm}$ may be expressed as

$$k_{xm}^2=k_{ym}^2-k^2.$$

Observe that the electromagnetic wave in the interaction space is damped along $k_i$ and propagated along $k_r$. The electromagnetic field in the $j$th slot of the sectorial resonator by (5a) and (5b), in corresponding notation, becomes

$$E_{yj}=-i\frac{kc^2}{\omega}B\{J_1(kr_a)+T_1(kr_b)N_1(kr_a)\}e^{i(k_{y0}jY-\omega t)}, \tag{7a}$$

$$B_{zj}=B\{J_0(kr_a)+T_1(kr_b)N_0(kr_a)\}e^{i(k_{y0}jY-\omega t)}, \tag{7b}$$

wherein the propagation vector $k_{y0}$ is related to $k_{ym}$ by the relation $k_{ym}=k_{y0}+2\pi m/Y$, in which $m$ is a positive or negative integer.

The boundary conditions imposed are

$$E_y=E_{yj}\left(jY-\frac{d}{2}\right)\leq y\leq\left(jY+\frac{d}{2}\right), \tag{8a}$$

$$E_y=0,\quad\text{for}\quad(jY-Y/2)\leq y\leq(jY-d/2), \tag{8b}$$

$$\text{for}\quad(jY+Y/2)\geq y\geq(jY+d/2),$$

$$B_z=B_{zj}\quad\text{at}\quad y=jY. \tag{8c}$$

The equality in (8c) is not strictly valid within

$$(jY-d/2)\leq y\leq(jY+d/2).$$

Although the field varies sinusoidally with $y$, it is small provided $d\ll Y$.

Hence, equating the electric and magnetic components of the electromagnetic fields, (6a, 6b,

---

[3] J. C. Slater, "The theory of the magnetron oscillator," M.I.T. Radiation Laboratory Report D-1-9, Office of Publications Board, Report PB-3777, p. 23, 1941. It may be procurable from the Department of Commerce, Washington 25, D. C. Slater's method, for the multi-rectangular cavity case described in his report, serves as a guide in our subsequent analysis.

7a, 7b) in the interaction space and sectorial cavities at the slots, and applying the associated boundary conditions (8a, 8b, 8c), the solution of (6a, 6b) and (7a, 7b) assumes the form

$$\sum_{m=0,\,\pm1\cdots}\frac{d}{Y}\frac{k}{k_{xm}}\left[\frac{J_1(kr_a)+T_1(kr_b)N_1(kr_a)}{J_0(kr_a)+T_1(kr_b)N_0(kr_a)}\right]\times\frac{1}{\tanh k_{xm}X}\frac{\sin\left(k_{y0}+\dfrac{2\pi m}{Y}\right)\dfrac{d}{2}}{\left(k_{y0}+\dfrac{2\pi m}{Y}\right)\dfrac{d}{2}}=1. \quad (9)$$

Recalling that

$$\Delta=kr_a-kr_a', \quad (10)$$

and introducing the contraction $Z_0(kr_a')$ in (4a),

$$J_0(kr_a')+T_1(kr_b)N_0(kr_a')=Z_0(kr_a')=0. \quad (11)$$

Observing that

$$Z_0'(kr_a)=-[J_1(kr_a)+T_1(kr_b)N_1(kr_a)].$$

Now, applying Taylor's expansion to (11), since $\Delta/k$ is sufficiently small, gives

$$Z_0(kr_a')=Z_0(kr_a-\Delta)$$
$$=Z_0(kr_a)-Z_0'(kr_a)\cdot\Delta+\cdots=0,$$

so that

$$\frac{J_1(kr_a)+T_1(kr_b)N_1(kr_a)}{J_0(kr_a)+T_1(kr_b)N_0(kr_a)}\simeq\frac{1}{\Delta}.$$

In addition it will be assumed that

$$k_{x0}\simeq k_{y0}=\frac{\pi}{Y}-\delta;\quad k_{x,-1}=\frac{\pi}{Y}+\delta$$

where $\delta$ is a small quantity,

$$\frac{\sin\left(k_{y0}+\dfrac{2\pi m}{Y}\right)\dfrac{d}{2}}{\left(k_{y0}+\dfrac{2\pi m}{Y}\right)\dfrac{d}{2}}\simeq1,$$

and

$$\tanh k_{xm}X\simeq1,$$

which implies in sequence that $k^2\ll(\pi/Y)^2$; $(Y/2(r_b-r_a))^2\ll1$, and $d$ and $X$ are sufficiently small and large, respectively, to justify these assumptions. With these assumptions, using but two terms of (9) in the region of large $k_{y0}$, Eq. (9) reduces to

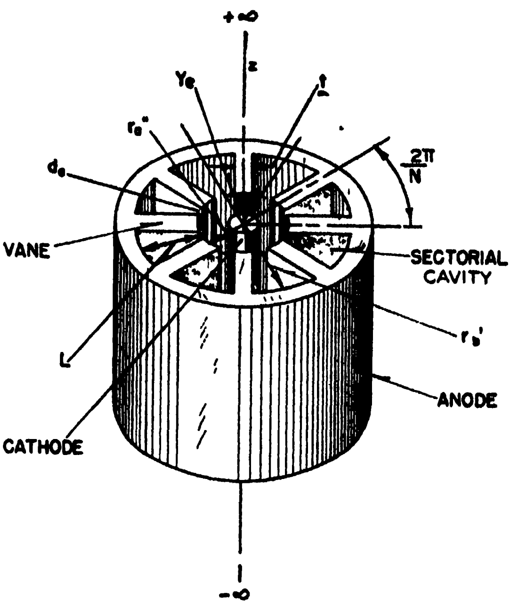


FIG. 3. A section of infinite length multi-sectorial cavity or vane magnetron.

$$\frac{d}{\pi}\frac{k}{\Delta}[(1-Y\delta/\pi)^{-1}+(1+Y\delta/\pi)^{-1}]=1, \quad (12)$$

which on applying the Binomial theorem, introducing $kL=kr_b-kr_a$, with the definition $(\pi/2)\Lambda(kL')=(kr_b-kr_a')$ where $L'=(r_b-r_a')$ and by virtue of (10), (12) reduces to the closed form for the wave-length

$$\lambda=\frac{1}{\Lambda(kL')}\left[4L+\frac{8d}{\pi}\{1+(\delta Y/\pi)^2\cdots\}\right], \quad (13)$$

which is valid for $k_{y0}(=\pi/Y-\delta)$ between $\pi/2Y$ to $\pi/Y$ for the linear case.

Now on the premise that only the region near the anode is of material consequence and that the velocity of propagation in the interaction space is unaltered, these results may be applied to the cylindrical magnetron, illustrated by Fig. 3. Incidentally, only discrete modes of oscillation are permitted in contrast to the linear case. Now, note that

$$\delta=\frac{\pi}{Y_e}-k_{y0},\quad Y_e=\frac{2\pi r_e}{N},$$

where $Y_e$ denotes the arc length between centers of adjacent cavities at $r_e''$.

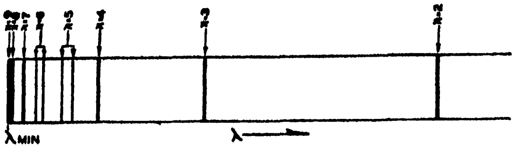


FIG. 4. An example of the fundamental multiplet band structure of modes measured on a Westinghouse 18-vane magnetron ($n_{\max}=N/2=9$; $n_{\min}=1$).

$k_{y0}r_a=n$
$=1, 2, \cdots N/2$ modes of oscillation for even $N$
$=1, 2, \cdots (N-1)/2$ modes of oscillation for odd $N$

and

$$\delta=\frac{\pi}{Y_e}\left(1-\frac{2n}{N}\right)$$

(12) for the cylindrical case, becomes

$$\lambda=\frac{1}{\Lambda(kL')}\left\{4L+\frac{8d_e}{\pi}\left[1+\left(1-\frac{2n}{N}\right)^2+\cdots\right]\right\}. \quad (14)$$

Since the $n=N/2$ or $\pi$ mode of oscillation is invariably sought, for reasons to be revealed presently, and the remaining modes of the fundamental multiplet receive consideration only from a competitive aspect, (14) may be simplified to read

$$\lambda=\frac{1}{\Lambda(kL')}\left(4L+\frac{8d_e}{\pi}\right), \quad (15)$$

which for the sake of brevity suffices in the subsequent discussions, although one can readily derive the complete spectrum of the fundamental multiplet from (14).

It is now appropriate to comment on the characteristics of the mode spectrum of a multicavity cylindrical magnetron in order to present a pictorial interpretation of the analytical results culminating in (14) for the vane magnetron.

The allowed modes of oscillation of the symmetrical cylindrical multicavity magnetron are grouped on a wave-length scale similar to band spectra. The longest wave-length band is referred to as the fundamental multiplet, its band structure or mode components are defined by Eq. (14) for the vane magnetron. It may be characterized by variations in $\phi$ of the electromagnetic field in the plane normal to the axis of the magnetron, for which edge effects are neglected. The modes of the band are characterized by phase differences between alternating potentials on the successive segments in the interaction space. Other bands of the spectrum have shorter wave-lengths, and their electromagnetic fields vary longitudinally as well. Each component of the fundamental multiplet has a different wave-length, although the differences are relatively small and depend upon the effective circuit constants of the cavity and the number of them. Figure 4 presents a pictorial

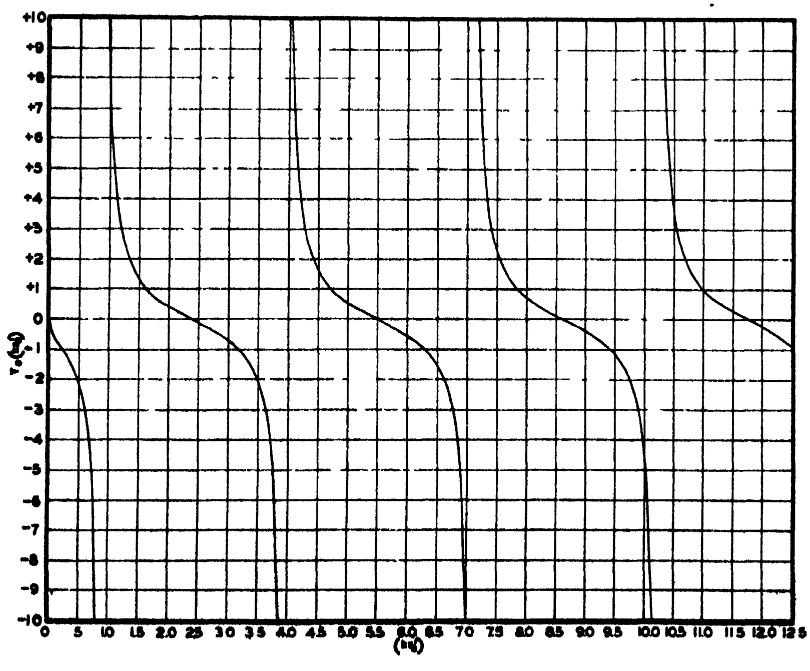
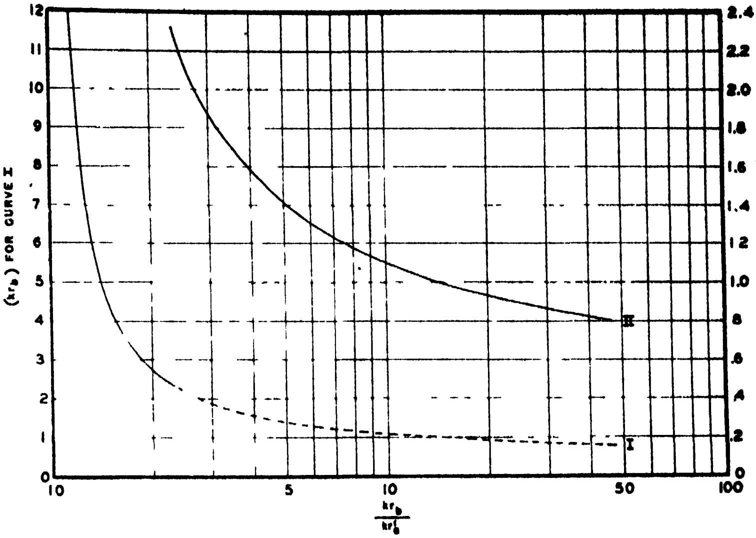


FIG. 5. Graph of the Bessel function, $T_0(kr_a')$, *versus* its argument $(kr_a')$. Values of $T_0(kr_a')$ for arguments less than 0.20 were evaluated from

$$T_0(kr_a')=-\frac{1}{\frac{2}{\pi}\ln\left[\frac{2}{\delta(kr_a')}\right]},$$

where $\delta=e^{\gamma}=1.781$, and $\gamma$ is Euler's constant.

FIG. 6. Solution of Eq. (4c), $T_0(kr_a')=T_1(kr_b)$. Curve II is a single dimensional magnification of the dashed portion of Curve I.



representation of the band structure or mode components of the fundamental multiplet from "cold" measurements for an unstrapped Westinghouse 18-vane magnetron. Modes 5 and 6 have been split into doublets by inherent unsymmetries. The relative intensities of the components in the fundamental multiplet are governed by the respective coupling between the resonators and the electronic system *via* the interaction space.

All except one of the modes in the fundamental multiplet for which the number of component resonators, $N$, of the magnetron is even, are doubly degenerate when the resonant system is completely symmetrical, whereas all the modes in the fundamental band, for which $N$ is odd, are doubly degenerate under the same conditions. The term doubly degenerate refers to the number of linearly independent modes, with the degenerate frequency, which in turn is defined as having more than one mode of oscillation though only one frequency. [It may be recalled however, that there are at least two exceptions for which certain unsymmetries do not result in removal of the degeneracy]. Such a split mode is known as a doublet and is characterized by two identical opposite traveling waves around the anode in the interaction space. For a symmetric magnetron with an even number of resonators there are, in the fundamental multiplet, $N-1$ modes with $n=1, 2, \cdots N/2$ separate wave-lengths, whereas for an odd $N$ there are $N-1$ modes though only $n=1, 2\cdots(N-1)/2$ separate wave-lengths. The non-degenerate mode is referred to as the $n=N/2$ or $\pi$ mode. Besides this non-degenerate mode there is an abnormal one, which is associated with the interaction space exclusively and is negligibly influenced by the multi-cavity geometry in a symmetrical multi-cavity system. This abnormal mode is referred to as the $n=0$ or zero mode and is usually well separated on a wave-length scale from the normal components of the fundamental multiplet. For these and other reasons its influence is negligible for symmetrical multi-cavity magnetrons. The electromagnetic field in the interaction space associated with these two modes is of the stationary type, though the zero-mode field is independent of the angular coordinate, whereas the $\pi$ mode field is periodic in this respect. The lower order modes of the fundamental multiplet of the infinite length symmetrical multi-cavity magnetron are ideally found at longer wave-lengths. The $\pi$ mode of the fundamental multiplet is characterized by non-degenerate property and highest conversion efficiency and for these reasons is invariably sought in practice.

In deriving Eq. (14) approximations which are explicitly or implicitly referred to were used. Hence, the present results must not be accepted unconditionally, though quantitatively they are not seriously in error.

Sometime after the completion of this work further efforts by others (in their unpublished

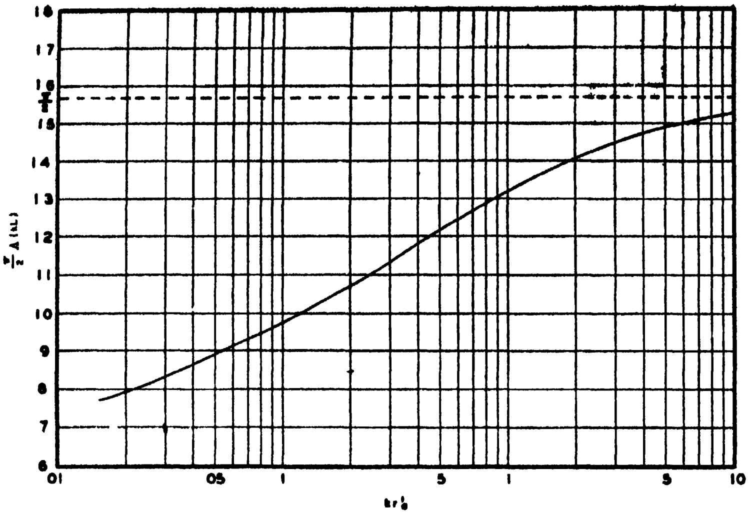


FIG. 7. A graph for the evaluation of the geometry of the sectorial cavity in terms of a given wave-length. It is derived from Fig. 5.

works),[4] concerning the infinite cylindrical case of the vane magnetron were able to be directed by the utilization of a method developed and applied by Clogston to the slot, and hole- and slot-type magnetrons in his unpublished thesis.[5] However, the influence of actual lengths of the end cavities and anode, which influence the order of the modes, their mutual separation, and consequently the wave-length of the $\pi$ mode are not explicitly included. Hence, predictions made in accordance with them, though more accurate in the ideal case than the results of the present method, are also inaccurate when applied to actual magnetrons by about the same percentage in some cases. Generally, however, they[4] are more accurate.

## III. APPLICATIONS

In order to utilize (13) it is necessary to first evaluate (4c). Figures 5 and 6 portray the results of this evaluation. Figure 7 is derivable from the results of Fig. 6. Incidentally, examination of Fig. 7 in conjunction with trigonometric expansions[6] for the components of Bessel functions (4) reveals that the transformation from the sectorial to the rectangular cavity, as $r_a$ and $r_b$ approach infinity together, provides that

$$kr_b - kr_a \to \pi/2$$

as the limit. So that

$$r_b - r_a \to \lambda/4,$$

which is the condition for the rectangular cavity.

Note that the solution of $kr_a'$ nearest to $kr_b$ is chosen by virtue of the mode considered in the sectorial cavity $\delta = 0$ in (13) for $n = N/2 = \pi$ mode.

In order to evaluate the radial vane length, $L$, or major cavity radius, $r_b'$, when the anode radius, $r_a''$, free space wave-length of a desired mode of oscillation, $\lambda_\pi$, vane thickness, $t$, and number of resonators, $N$, are known, the following procedure is convenient. Refer to Fig. 3 for identification of symbols.

Given data:

wave-length, $\lambda_\pi = 9.46$ cm

$$\left(k = 0.664 \frac{1}{\text{cm}},\right)$$

number of sectorial cavities, $N = 18$,
vane thickness, $t = 0.263$ cm,
anode diameter, $2r_a'' = 3.46$ cm,
effective anode length $= 3.36$ cm,
cathode diameter, $2r_c = 1.100$ cm,
cathode to anode ratio, $r_c/r_a'' = 0.318$.

Determine $L$ or $r_b'$ by the following procedure:

the coupled cavity radius,

$$r_a = r_a'' - \frac{t}{2 \sin \pi/N} = 0.972 \text{ cm},$$

---

[4] W. Altar, Westinghouse Research Report SR-133, September 15, 1942, and later by H. Goldstein, M.I.T., Ph.D. thesis of 1943.

[5] A. M. Clogston, M.I.T., Ph.D. thesis of 1941.

[6] Jahnka-Emde, *Tables of Functions* (B. G. Teubner, Leipzig and Berlin, 1938), p. 138.

equivalent slot width,

$$d_e=\frac{\pi 2r_a''}{N}-t=0.441\text{ cm},$$

from relation (12), for the $\pi$ mode $\delta=0$, so

$$\Delta=\frac{2d_e k}{\pi}=0.1867,$$

from Fig. 7, for $kr_a'=kr_a-\Delta=0.459$; $(\pi/2)\Delta(kL')=(kr_b-kr_a')=1.20$,
hence $kL=(kr_b-kr_a')-\Delta\simeq 1.013$ cm,
so that $L=1.525$ cm
and $2r_b'=6.51$ cm,
which differs by about 6 percent from its experimental value on a magnetron (Westinghouse VA-1) possessing these dimensions.

In order to evaluate the wave-length for the $\pi$ mode of oscillation when the anode radius, major cavity radius, vane thickness, and number of resonators are known, it is merely necessary to evaluate the ratio $(kr_b/kr_a')$ and $kr_b$ with the aid of Fig. 6, as illustrated by the following results for the previous Westinghouse magnetron. Refer to Fig. 3 for identification of symbols.

Given data:

number of sectorial cavities, $N=18$,
vane thickness, $t=0.263$ cm,
anode diameter, $2r_a''=3.46$ cm,
effective shell diameter, $2r_b'=6.90$ cm,
effective anode length $=3.36$ cm,
cathode diameter, $2r_c=1.100$ cm,
cathode to anode ratio, $r_c/r_a''=0.318$.

Determine $\pi$ mode wave-length, $\lambda_\pi$, by the following procedure:

$$d_e\simeq\frac{\pi 2r_a''}{N}-t=0.341\text{ cm}$$

$$r_b=r_b'-\frac{t}{2\sin\pi/N}=2.693\text{ cm}$$

$$r_a=r_a''-t/2\sin\pi/N=0.973\text{ cm}$$

$$r_a'=r_a-2d_e/\pi,\ \{\text{from (10) and (12) }\delta=0\text{ for }n=\pi\text{ mode}\},\ =0.756\text{ cm}$$

$$r_b/r_a'=3.54$$

$$(kr_b)\text{ from Fig. 6}=1.70$$

$$\lambda_\pi=2\pi r_b/(kr_b)=9.95\text{ cm}$$

percent difference $\simeq$5 percent from operating wave-length of 9.46 cm for $n=9$ or $\pi$ mode. (Westinghouse VA-1.)

Because of assumptions, such as infinite anode length, linear approximation, etc., inherent in the present method it is not surprising to be confronted with differences of the order of 10 percent more or less between the experimental and theoretical values of $\pi$ mode wave-length. For instance, the length of the anode and end cavity

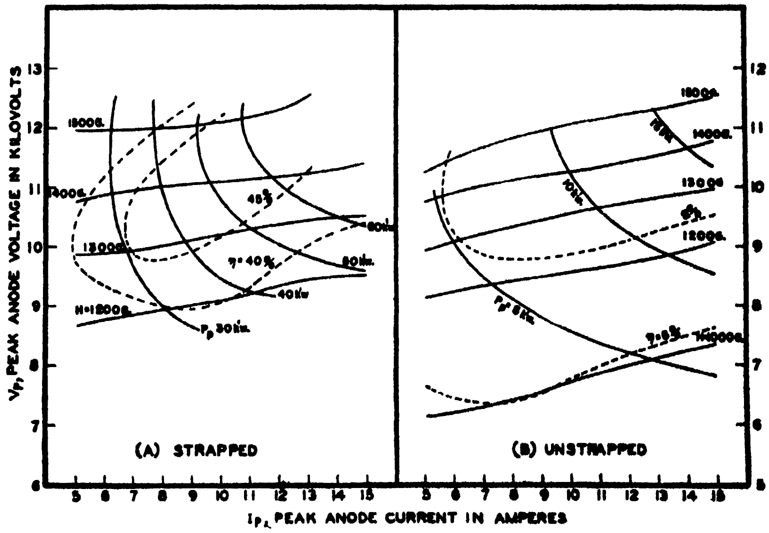


FIG. 8. Performance charts of Westinghouse 8-vane magnetrons, WX-3256, illustrated in Fig. 9. The magnetrons are terminated by a 50-ohm matched coaxial line. The operating unstrapped and strapped wave-lengths are 8.0 and 9.6 cm, respectively.

in terms of free space wave-length will have more or less of an influence on the order and separation of the modes in the multiplet and obviously on the wave-length of the $\pi$ mode.[7] The method, nevertheless, provides reasonable approximate data for a preliminary design, which can then be corrected accordingly by "cold" test.

After examining "cold" test data of a number of similar 8-vane unstrapped 10-centimeter magnetrons constructed with the aid of the preceeding theory, it was discovered that for end cavity and effective anode lengths of the order of $\lambda/10$ and $\lambda/4$ of free space wave-length, respectively, that the perimeter of the sectorial cavity normal to the axis of the magnetron, was nearly equal to half the free space $\pi$-mode wave-length. Since the end cavity and anode lengths have a more or less influence on the order, mode separation, and wave-lengths of the modes of the fundamental multiplet in a practical magnetron, this perimeter index is useful only as a guide in deriving the magnitude of the $\pi$-mode wave-length for a vane magnetron

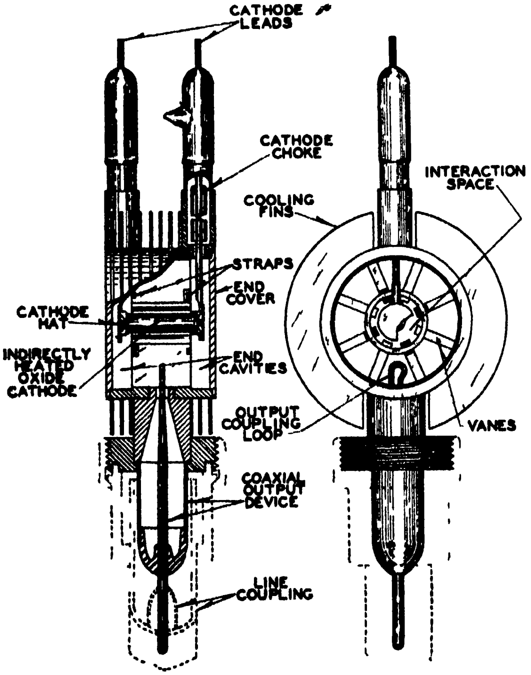


FIG. 9. Westinghouse vane magnetron, WX-3256.

## IV. APPLICATION OF STRAPPING TO THE RESONATOR SYSTEM

Because of the inherent assumptions in the theoretical methods for the determination of mode-frequency spectrum of a multi-cavity magnetron, varying degrees of error result in application so that one often uses the theoretically derived values only as a guide in obtaining a magnetron-resonator system possessing mode-frequency spectrum in the desired wave-length range, for subsequent "cold" tests. From such tests are derived the actual mode spectrum; and if undesirable for one reason or other the resonator system is modified in order to achieve the desired mode separation (about 5 percent or more) and distortion of competing modes, for efficient operation. When the modes are too close together, ($\ll$5 percent say) as is the case with symmetrical multi-cavity magnetrons (for $N \gg 2$) without straps, the competing modes cause the efficiency of the desired mode to be low, as illustrated by Fig. 8. In order to achieve the requisite mode separation between the desired mode (usually the $\pi$ mode since it is non-degenerate and associated with the highest theoretical conversion efficiency) and competing ones in the fundamental spectrum of a symmetrical magnetron, in contrast to the "rising sun" type, strapping is resorted to. Strapping is a mechanical means of joining segments of like phase, usually at one or both ends of an anode, adjacent to the interaction space, in such manner as to increase the normal coupling between resonators and hence the effective capacity of the resonator system. Strapping usually distorts the competing mode patterns so that their interaction with the space-charge field at the slot is poor. The $\pi$ mode is not significantly distorted by proper strapping, so that efficient electronic interaction ensues for this mode. The coupling between cavities for a particular mode may be selective with straps and thereby further increase mode separation.

Essentially two types of strapping are utilized. The symmetrical type of strapping, such as the "echelon,"[7] separates the components of the fundamental multiplet but actually does not split the degenerate modes, whereas the unsymmetrical type of strapping, such as the "Y–B,"[7] which adequately distorts the competing modes, only slightly influences the $\pi$-mode pattern. The type

---

[7] J. C. Slater, M.I.T. Radiation Laboratory Report No. 182, "Resonant modes of the magnetron." It may be procurable from the Office of Publications Board, previously referred to.

TABLE I. Magnetron scaling.

| Ratio of initial to final parameter | General scaling $Q_L'/Q_{L=1}\ r_c'/r_c=\left(\frac{\sigma}{\sigma'}\right)\left(\frac{r_a'}{r_a}\right)$; $\sigma_{\text{opt}}=\frac{1+4/N}{1-4/N}$ (see reference 3 and Section VI) | Restricted scaling $\frac{N'}{N}=1;\ \frac{Q_L'}{Q_L}=1,\ \frac{\pi'}{\pi}=1;\ \frac{\sigma'}{\sigma}=1$ |
|---|---|---|
| $\frac{H'}{H}$ | $\frac{\lambda}{\lambda'}$ | $\frac{\lambda}{\lambda'}$ |
| $\frac{V'}{V}$ | $(r_a'/r_a)^2(N/N')(\lambda/\lambda')^2\left[\frac{1-\frac{1}{\sigma'}}{1-\frac{1}{\sigma}}\right]$ | $(r'/r)^2(\lambda/\lambda')^2$ |
| $\frac{I'}{I}$ | $(L'/L)(r_a'/r_a)^2(N/N')(\lambda/\lambda')^3$ | $(L'/L)(r'/r)^2(\lambda/\lambda')^3$ |
| $\frac{J_c'}{J_c}$ | $(r_a'/r_a)(N/N')(\lambda/\lambda')^3(\sigma'/\sigma)$ | $(r'/r)(\lambda/\lambda')^3$ |
| $\frac{P'}{P}$ | $(L'/L)(\lambda/\lambda')^5(r_a'/r_a)^4(N/N')^2\left[\frac{1-\frac{1}{\sigma'}}{1-\frac{1}{\sigma}}\right]$ | $(L'/L)(r'/r)^4(\lambda/\lambda')^5$ |

of assymmetry and degree of coupling or "tightness" of strapping will govern the mode separation and number of degenerate modes which are split by unsymmetrical strapping.

The use of unsymmetrical straps provides one of several necessary conditions for efficient operation in the $\pi$ mode. Figure 8(a) illustrates a typical improvement in performance over that of Fig. 8(b) for the same 10-centimeter, 8-vane magnetron after straps and strap-breaks are incorporated, as illustrated by Fig. 9. "Single recess ring" straps at both ends of the anode are illustrated. The cathode "hat" illustrated, serves to establish an axial component of electric field in order to reduce the number of electrons which would otherwise leave the interaction space and bombard the end covers.

The ratio between the unstrapped, $\lambda_{\pi U}$, and strapped, $\lambda_{\pi S}$, wave-lengths may be expressed

$$\left(\frac{\lambda_{\pi U}}{\lambda_{\pi S}}\right)^2=\frac{1}{1+\dfrac{C_S}{C_T}},$$

in terms of $C_S$ and $C_T$, which denote the total lumped effective values of strap to vane- and resonator-system capacities, including mutuals. The resonator-system capacity is a product of the number of resonators and the effective lumped capacity of each cavity, less straps. The strap capacity will obviously vary with the number and type of straps used and the geometry at the strap-vane intersection. Both $C_T$ and $C_S$ may be derived either by simple electrostatic approximations, by means of models in the electrolytic tank, or by the more elaborate conformal transformations. Practical results are available.[8]

To increase further effectiveness of straps, strap-breaks over the segments are incorporated at judicious places with respect to the load coupling in order to distort the unwanted mode by proper phasing and reduce the tendencies to mode shifts during load variations. With proper type of strapping and number and orientation of strap-breaks, relatively stable competing-mode free operation may be realized.

Since the ratio of equivalent effective capacity to inductance of the cavity (which is a product of the reciprocal of the number of cavities and the effective inductance of one of them) of a magnetron-resonator system is altered by strapping, the frequency pulling characteristics are also

---

[8] E. Everhart, M.I.T. Radiation Laboratory Report 223, "Magnetron strapping, wave-length calculations for strapped magnetrons." Office of Publications Board Report, PB-2850, 1943. W. V. Smith, M.I.T. Radiation Laboratory Report 226, Office of Publications Board Report, PB-8327, 1943, "Practical considerations of magnetron design." These reports may be procurable from the Office of Publications Board previously referred to.

Table II. Restricted scaling in wave-length and voltage.

| Ratio of initial to final parameter | Wave-length scaling (shape, invariant) $\lambda'/\lambda = r'/r$, $L'/L = \lambda'/\lambda$ | Voltage scaling (wave-length invariant) $\lambda'/\lambda = 1$, $L'/L = 1$ |
|---|---|---|
| $H'/H$ | $\lambda/\lambda'$ | 1 |
| $V'/V$ | 1 | $(r'/r)^2$ |
| $I'/I$ | 1 | $(r'/r)^2$ |
| $J'/J$ | $(\lambda/\lambda')^2$ | $r'/r$ |
| $P'/P$ | 1 | $(r'/r)^4$ |

altered. Of course, altering the anode length and end cavity length of a symmetrical magnetron with or without straps will influence the mode separation, mode wave-length, and order or distribution between components of the fundamental multiplet. For instance, by altering the anode length of an unstrapped anode, of given end cavity geometry, the order between the components may be completely reversed and consequently wave-lengths changed.

## V. APPLICATION OF DIMENSIONAL SCALING[9]

After incorporation into the magnetron of all the necessary features known to be desirable and after obtaining performance characteristics, a suitable prototype may be obtained which may be dimensionally scaled to other wave-lengths, voltages, powers, etc., while the original efficiency is retained during the transformation. Scaling of vane magnetrons requires that the geometry of the straps, its vane recess, and position on the vane be scaled so as to maintain the strap in the original relative position on the vane. Scaling of a prototype is a very profitable design procedure since it avoids the expensive and time consuming development which would otherwise be required. Since these scaling relations are very useful and appropriate to the present subject, they will be briefly considered herein, though without derivation. Approximate scaling relations have been derived,[10] on the basis of Newton's and Maxwell's differential equations applied to the electrons in the interaction space of the magnetron, and provide conditions for which two magnetrons, differing by some scaling factor, possesses the same efficiency or equivalent operation. These conditions of anode voltage, $V$, and magnetic field, $H$, for constant ratio of anode to cathode radii, $\sigma$, mode of oscillation, $n$, or number of cavities, $N$, and loaded, $Q$, $Q_L$, may be expressed as the invariant magnetron scaling relations,

$$H\lambda,\ V\cdot\left(\frac{\lambda}{r_a}\right)^2;\ \frac{I\lambda^3}{r_a{}^2L},$$

in which $V$ and $H$ are frequently expressed in kilovolts and Gauss, respectively; anode current, $I$, is expressed in amperes, anode length, $L$, anode radius, $r_a$, and wave-length, $\lambda$, in centimeters. This restricted type of scaling is illustrated in terms of the performance chart parameters, for constant $\sigma$, $n$, $N$, and $Q_L$, in Table I. Note that the scaling is valid for either input or output values of anode voltage, $V$, and power, $P$. The current density, $J$, is applicable at either the cathode or anode.

Table I also illustrates the general scaling relations, in which the restraints on the restricted type of scaling have been removed. Note that for the general scaling the current density is applicable only at the cathode. Another relation can be derived for the anode. Very often it is desirable to wave-length, voltage, or power scale a desired operating point of a magnetron under the conditions of the restricted case. For this purpose, the corresponding restraints on the shape and wave-length of the magnetron are inserted into the expressions for the restricted ($N$ fixed) case. The result of which is illustrated by Table II. For both the restricted and general types of scaling, end effects are neglected such that the anode current, $I$, per unit length is independent of length.

With the aid of the scaling relations illustrated by Tables I and II it is possible to scale a selected operating point on the performance chart of a magnetron, as the prototype, to another point in terms of the operational parameters. This process may or may not involve a change of the number of cavities,[11] depending upon whether the general or restricted forms of scaling are applied. Though very often it is only necessary to apply the

---

[9] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), Section 9, p. 488.

[10] A. M. Clogston, M.I.T. Radiation Laboratory Report 52. This was later extended by H. C. Hu in Radiation Laboratory Report 1043-3. These reports may be procurable from the Office of Publications Board, previously referred to.

[11] When a large change in the number of cavities, $N$, is contemplated, in general scaling, the error may be significant, so that one ought to consult reference 10 by H. C. Hu for details.

restricted type of scaling in the forms illustrated by Table II.

The application of the scaling processes illustrated by Table I is readily acquired. For example, consider the often used restricted scaling process in the form illustrated by Table II, which is to be applied to a prototype in order to permit operation at another wave-length and power output. The first step is to make a wave-length scale by altering the effective magnetron's dimensions linearly, in proportion to the new to the old wave-length, in accordance with column two of Table II Then voltage scale by evaluating the anode radius of the interaction space from a knowledge of the desired power by the appropriate relations given under the restricted voltage scaling column. Obviously, one must use the physical and operational parameters derived from wave-length scaling for voltage scaling. This two-step scaling process is convenient, since it affords an inspection of the change in parameters from the prototype to the fully scaled magnetron. Sometimes both power and wave-length must be increased so much above that of the prototype that the consequence of increased physical size may be undesirable; then the prototype may be only wave-length scaled to the desired region according to Table II. After this the desired power may or may not be achieved with undesirable characteristics by mere extrapolation of the applied fields above the chosen operating point of the prototype. In scaling or extrapolation involving a change of the cathode-current density one should, for economic and stability considerations, maintain this as near its stable value as possible. The extent to which simple extrapolation of this sort is practical is usually unknown in advance of tests, because of various implicit properties of the magnetron. Obviously, the performance characteristics of such a "new" magnetron, if operable, are no longer those of the prototype since the above invariance conditions are not satisfied by such an extrapolation.

## VI. RELATED TOPICS AND PRACTICAL CONSIDERATIONS

The distribution of the electric component of the electromagnetic field in the sectorial cavity may be plotted from (5b) as a function of $(r-r_a)/(r-r_b)$, with $r_b/r_a$ as a parameter.

Using (5a) and 5(b), together with the boundary conditions (4a) and (4b), enables the evaluation of the total stored energy per unit length and hence the $Q$.

Practically, for symmetrical magnetrons having $N=6$ to 16 inclusive, the reduced cathode radius, $1/\sigma$, varies from 0 to 17 percent greater than Slater's[3] theoretical optimum value, $(=(1-4/N)/(1+4/N))$, and for $N=18$ this reduced cathode radius amounts to about 5 percent less than this theoretical optimum value. However, the reduced cathode radius may radically deviate from its theoretical optimum value when considerations related to oscillation build-up time and maximum $\pi$-mode (applied) voltage or current in pulsed and C.W. magnetrons are of paramount importance. For example, the 5J26 magnetron possesses a reduced cathode radius amounting to 64 percent greater than the theoretical optimum value, and this at 48 percent over-all efficiency, which incidentally is a product of 58 percent electronic and 82 percent circuit efficiencies.

In voltage scaling symmetrical, strapped magnetrons one must consider the alteration in mode separation, $(\lambda_{n_\pi}-\lambda_{n_\pi\pm1})$, by virtue of the alteration in strap lengths. Since there exists an optimum mode separation (except possibly for rising sun structures with large $N$) for the attainment of the maximum $\pi$-mode input voltage (or current), aside from any further curtailment under pulsed conditions by virtue of inadequate oscillation build-up time, one should determine whether or not the resultant mode separation will seriously curtail the desired $\pi$-mode input voltage or current, under either C.W. or pulsed conditions, after voltage scaling.

# Gyroscopic Effects in a Vibratory System*

G. S. Bennett**
*Naval Research Laboratory, Washington, D. C.*

(Received June 18, 1947)

An analytical investigation is made to determine the influence on a vibrating elastic system of a high speed rotating element. A fundamental condition for the presence of any gyroscopic effect is that the vibration be such as to produce a tilt of the spin vector of the rotor. Equations are derived showing (a) that if precession be prevented, no gyroscopic effects are present, and (b), that if precession be permitted, the gyroscopic action contributes a force opposed to the tilt, and proportional to the angle of tilt, in effect increasing the stiffness of the elastic system. Calculations based on an arbitrarily chosen rotor intended to approximate a typical small motor armature indicate that this additive spring force is small, and that the increase in natural frequency caused by the added stiffness is exceedingly small—on the order of 0.1 percent—in the case where the precessional motion is not restrained

IN many cases, especially aboard aircraft, equipments subjected to vibration have as component parts dynamotors or other rotating elements. It is the purpose of this investigation to determine whether gyroscopic effects caused by rotating elements exert any marked influence on the vibration characteristics of the equipment.

The fundamental condition necessary for gyroscopic action to appear is that a torque be applied tending to change the direction of the axis of rotation of a spinning body. A pure translation will produce no such effects. Assuming such a torque, and using Euler's angles (see Fig. 1) as coordinates for a rigid body in space, we can write Euler's equations for the applied torques about the three principal axes as[1]

$$M_1=A\omega_1-(B-C)\omega_2\omega_3, \quad (1a)$$

$$M_2=B\omega_2-(C-A)\omega_3\omega_1, \quad (1b)$$

$$M_3=C\omega_3-(A-B)\omega_1\omega_2, \quad (1c)$$

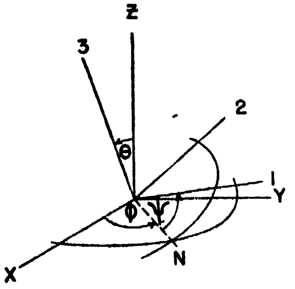


FIG. 1. Euler's angles as coordinates of a rigid body in space.

where $A$, $B$, and $C$ are the moments of inertia about the principal axes 1, 2, and 3, respectively.

Consider the case of a rigid object symmetrical about one axis, and identical with respect to the other two, such as a cylinder rotating about its axis of figure, or a motor armature. Specifying the principal axes and moments of inertia as in Fig. 2, it is apparent that $A=B$. Writing $(B-C)=-(C-A)=H$, and stipulating that $M_3=0$, by use of the equations

$$\omega_1=\dot{\theta}\cos\psi+\phi\sin\theta\sin\psi, \quad (2a)$$

$$\omega_2=\dot{\theta}\sin\psi+\phi\sin\theta\cos\psi, \quad (2b)$$

$$\omega_3=\psi+\phi\cos\theta \quad (2c)$$

substituted in Eqs. (1), we easily find

$$M_1=\cos\psi[A\theta+\phi\psi C\sin\theta-H\phi^2\sin\theta\cos\theta] +\sin\psi[A(\partial^2\phi/\partial t^2)\sin\theta-C\theta\psi +(A+H)\phi\theta\cos\theta] \quad (3a)$$

$$M_2=\cos\psi[B(\partial^2\phi/\partial t^2)\sin\theta-C\theta\psi +(B+H)\phi\theta\cos\theta]-\sin\psi[B\ddot{\theta}+\phi\psi C\sin\theta -H\phi^2\sin\theta\cos\theta] \quad (3b)$$

$$0=C\psi+C(\partial^2\phi/\partial t^2)\cos\theta-C\phi\theta\sin\theta]. \quad (3c)$$

From Eqs. (1c) and (2c), in place of Eq. (3c), we may write

$$\omega_3=K=\psi+\phi\cos\theta. \quad (4)$$

Examination of Eqs. (3a) and (3b) shows immediately that the torques about the moving principal axes 1 and 2 may be considered as the projections of two torques not rotating with the principal axes; one lying along the nodal line $N$ and designated as $P$, the other perpendicular to this line but in the 1, 2 plane and designated as

---

* Paper presented at the 277th meeting of the American Physical Society.

** Now at Michigan State College, East Lansing, Michigan.

[1] Slater and Frank, *Introduction to Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1933), first edition, Chapter 10.

$Q$. Thus,

$$P=A\ddot{\theta}+\dot{\phi}\sin\theta(CK-A\dot{\phi}\cos\theta), \tag{5a}$$

$$Q=B(\partial^2\phi/\partial t^2)\sin\theta-\dot{\theta}(CK-2B\dot{\phi}\cos\theta). \tag{5b}$$

The first term of Eq. (5a) is recognized as the usual Newtonian torque, while the second term is gyroscopic in nature. The first term of Eq. (5b) is, similarly, the Newtonian torque about the precession axis making an angle $(\pi/2-\theta)$ with the 1, 2 plane caused by a torque lying in that plane. Again, the second term of Eq. (5b) is gyroscopic in nature.

There are two cases to be considered; (I) precession is prevented by suitable restraints, and (II) the torque $Q$ is zero (precession not restrained).

For case I, Eqs. (5) reduce to

$$P=A\ddot{\theta}, \tag{6a}$$

$$Q=-CK\dot{\theta}, \tag{6b}$$

where $Q$ is then the torque exerted by the restraint in order to prevent precession, and $P$ the torque which produces tilting—the Newtonian torque.

For case II, rewriting Eqs. (5) with the aid of Eq. (4), and setting $Q=0$,

$$P=A\ddot{\theta}+(K-\dot{\psi})\tan\theta[CK-A(K-\dot{\psi})], \tag{7a}$$

$$0=B(\partial^2\phi/\partial t^2)\sin\theta-C\dot{\theta}(2K-\dot{\psi})+2B\dot{\theta}(\dot{\psi}-K). \tag{7b}$$

Substituting $\theta$ for $\tan\theta$ (for small $\theta$) and letting $\dot{\psi}$ be constant, we can write

$$P=A\ddot{\theta}+R\theta, \tag{8a}$$

$$0=B(\partial^2\phi/\partial t^2)\sin\theta-S\dot{\theta}. \tag{8b}$$

Equation (8a) shows that the gyroscopic effect is that of adding a spring stiffness to the inertia torque. If the system be mounted on springs in the $\theta$ direction, the effect is that of raising the natural frequency. Equation (8b) shows that the torque producing precession is proportional to the velocity of tilt.

Let us now examine the relative importance of the constants $R$ and $S$. Equations (7) and (8) give

$$R=CK(K-\dot{\psi})-A(K-\dot{\psi})^2, \tag{9a}$$

$$S=C(2K-\dot{\psi})-2A(K-\dot{\psi}). \tag{9b}$$

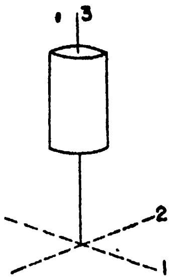


FIG. 2. Arbitrarily chosen rotor. 1, 2, and 3 are principal axes about which the moments of inertia $A$, $B$, and $C$ are taken, respectively. Rotor is taken to be 5 inches long, 2 inches in diameter, and with the lower end 5 inches from the 1, 2, plane. The weight of the rotor is 5 lbs. and the spin velocity chosen as 6000 r.p.m.

Now in general, for systems of the sort here considered, the resultant angular velocity $K$ will be only slightly greater than the spin velocity $\dot{\psi}$, so that $(K-\dot{\psi})$ is small.

From Eq. (9a), $R$ is a quadratic function of $(K-\dot{\psi})$, and is zero when $(K-\dot{\psi})=0$, and when $(K-\dot{\psi})=CK/A$, or $(K-\dot{\psi})/K=C/A$. Between these limits, $R$ reaches a maximum at $(K-\dot{\psi})/K=C/2A$, and for values outside these limits $R$ is negative. Now a negative value of $R$ is physically absurd, as this would indicate a gyroscopically produced torque tending to aid the external torque, which is contrary to all experience. Thus $(K-\dot{\psi})/K$ must lie between $0$ and $C/A$.

For a rough estimate as to the magnitude of $R$, let us assume the rigid body as is shown in Fig. 2 with the axes of rotation as shown. This is thought to be a fair representation of a typical motor armature. Assuming the dimensions and weight shown, and a motor speed of 6000 r.p.m., the maximum value of $(K-\dot{\psi})/K=C/A=0.019$ and the value of $(K-\dot{\psi})/K$ for maximum $R$ is $C/2A=0.0095$. Putting these values in Eq. (9a), the maximum value $R$ can assume is found to be 10,500 lb. in./sec., which gives for a natural frequency about the fixed point (in the $\theta$ plane) approximately 0.95 c.p.s. This is much lower than the usual run of vibration frequencies, implying that the added apparent spring stiffness is small compared to the actual springs present.

$S$ can be rewritten as

$$S=CK-(K-\dot{\psi})(2A-C), \tag{10}$$

which is a linear function of $(K-\dot{\psi})$. Again, $S$

must be always positive, or the limiting value of $(K-\psi)/K=C/(2A-C)$, which is somewhat larger than the value found above giving the maximum value of $R$. As $C$ becomes small compared to $A$, the value of $(K-\psi)/K$ approaches the value for maximum $R$, and as $C$ approaches $A$, the value of $(K-\psi)/K$ approaches the value found above as an upper limit. It is to be noted that $S$ is a direct function of $K$, the resultant angular velocity—as the spin velocity increases, the torque-producing precession increases. For the particular values here chosen, (for maximum $R$) $S$ is 35.5 in lb $^2$/sec.$^2$.

The rotor might well be mounted on springs such that the stiffness would be four-hundred times the apparent stiffness calculated here. For the rotor alone this would mean a natural frequency of 19 c.p.s., which is not uncommon. The effect of added inertia would be to reduce both natural frequencies (real and gyroscopic) without affecting the spring stiffness. In this case then, the gyroscopic effect will be to raise the natural frequency in the ratio

$$\frac{f_{g+s}}{f_s}=\left(\frac{401}{400}\right)^{\frac{1}{2}}=1.0012,$$

and the change is on the order of 0.1 percent.

# Frequency Stabilization of Microwave Oscillators by Spectrum Lines*

William V. Smith, José L García de Quevedo, R L Carter, and W S Bennett
*Department of Physics, Duke University, Durham, North Carolina*

(Received June 25, 1947)

A 2K50 microwave oscillator has been electronically coupled to the $NH_3$ 3,3 rotational spectrum line at 23,870 megacycles in such a manner as to maintain the frequency of the oscillator at the spectrum-line frequency The theory of the "spectrum-line discriminator" designed to produce this frequency stabilization is developed.

## INTRODUCTION

THE possibility of stabilizing the frequency of a microwave oscillator by means of a spectrum line has already been mentioned in the literature.[1] The present article deals with one form of "spectrum-line stabilizer" which has been used to tie the frequency of a 2K50 klystron to the $NH_3$ 3,3 spectrum line at 23,870 megacycles. It is believed that this is the first time any oscillator has been thus "tied to a spectrum line" although variations in the absorption properties of gases have been previously used to modify the properties of cavity stabilizers.[2] Since the first laboratory model of the stabilizer is relatively simple and reliable, it is to be anticipated that engineered versions of the stabilizer will be of considerable importance not only as frequency standards substantially independent of temperature, but also as frequency controlling elements of communications devices, since transmitter and receiver can be set unerringly on the same frequency merely by tying each to the same spectrum line

## THEORY

In developing the theory of this spectrum-line stabilizer, frequent reference will be made to Pound's article on the stabilization of microwave oscillators,[1] as the present stabilizer design is in some features similar to his direct-current cavity stabilizer circuit. The comparable elements of the two stabilizers, i.e., the discriminators, are shown in Figs. 1a and 1b. The operation of both circuits is, essentially, that an input signal of variable frequency is reflected from the terminations of the two tee sections, finally being rectified by crystals $A$ and $B$. The reflections are frequency sensitive, particularly the reflection from arm 2 of tee $A$, which can be so designed that the rectified output of crystals $A$ and $B$, suitably added, is a linear function of frequency near the resonant frequency $\nu_0$ of the element terminating arm 2. In general, the output is adjusted to be zero at $\nu_0$, and is amplified and applied to the re-

---

* The research described in this report was supported by Contract No. W-28-099-ac-125 with the Army Air Forces, Watson Laboratories, Air Materiel Command

[1] R. V. Pound, Rev. Sci. Inst. 17, 490 (1946).

[2] A. Roberts, Y. Beers, and A. G. Hill, Phys. Rev. 70, 112 (1946).

flector electrode of a microwave oscillator. Since the frequency of oscillation of the oscillator is a linear function of the voltage applied to the reflector electrode, it is possible to feed back the output of the discriminator circuit in such a phase as to correct for any tendency of the oscillator to drift away from the resonant frequency $\nu_0$ of the discriminator.

The only apparent difference between the two discriminators consists in replacing Pound's cavity at $x-\lambda_g/8$ by a short circuit and inserting symmetrical windows $C$ and $D$ in the tee near its junctions, so that one or both of the lengths of guide can be evacuated and filled with gas at a low pressure. (These windows may conveniently be made of mica, and if window $D$ is reflectionless and arm 1 is filled with air, window $C$ can be omitted.) If, then, the gas in arm 2 is resonant at a frequency $\nu_0$, it can be shown that the combined reflections from arms 1 and 2 are the same as for a cavity of frequency $\nu_0$ terminating arm 2, when no gas is present in the arm. Physically, the phenomena occurring in the case of the spectrum-line discriminator are different from those in the cavity discriminator. Energy is absorbed all along arm 2 in the first case, and only in the terminating cavity in the second case. Also, for the spectrum-line case arms 1 and 2 must be long enough so that appreciable absorption takes place at $\nu_0$, but not so long that the incident wave is completely absorbed at $\nu_0$, as then complete absorption may also take place at frequencies in the immediate vicinity of $\nu_0$. For the cavity discriminator, on the other hand, the optimum condition is a match at $\nu_0$, i.e., complete absorption of incident energy. Finally, it should be noted that in the spectrum-line case the two arms differ by $\lambda_g/8$ in electrical length, which is not necessarily the same as the physical length as the dielectric constant (at a frequency well removed from $\nu_0$) may differ in the two arms.

The mathematical analysis will be undertaken for arms 1 and 2 of equal electrical lengths $l$, and the reciprocal of the difference in input impedances to the two arms will be shown to be the same as the input admittance of a cavity. It then follows that if the lengths differ by $\lambda_g/8$, the electrical properties of Fig. 1b will be those of the cavity discriminator of Fig. 1a. Expressing the input impedances $Z_1$ and $Z_2$ to arms 1 and 2 in terms of the characteristic impedance $Z_0$ of the guide, and ignoring the small change in $Z_0$ caused by the gas in comparison with the change in propagation constant, then

$$\begin{aligned} Z_1 &= Z_0 \tanh \gamma_0 l, \\ Z_2 &= Z_0 \tanh \gamma l, \end{aligned} \tag{1}$$

where $\gamma_0$ and $\gamma$ are the propagation constants in vacuum and in gas.

If $\alpha_0$ is the attenuation of the guide walls in nepers per cm, and $\epsilon_1$ and $\epsilon_2$ are the real and imaginary parts of the dielectric constant of the gas,

$$\begin{aligned} \gamma_0 &= \alpha_0 + \frac{2\pi}{\lambda_g} j, \\ \gamma &= \gamma_0 + \frac{\pi}{\lambda_g}[j(\epsilon_1 - 1) + \epsilon_2]. \end{aligned} \tag{2}$$

Also, according to Van Vleck,[3] if the line breadth $\Delta\nu$ is caused solely by collision broadening

$$\begin{aligned} \epsilon_1 - 1 &= \frac{A\nu_0}{2} \frac{\nu_0 - \nu}{\Delta\nu^2 + (\nu - \nu_0)^2}, \\ \epsilon_2 &= \frac{A\nu_0}{2} \frac{\Delta\nu}{\Delta\nu^2 + (\nu - \nu_0)^2}, \end{aligned} \tag{3}$$

where $A$ is a constant.

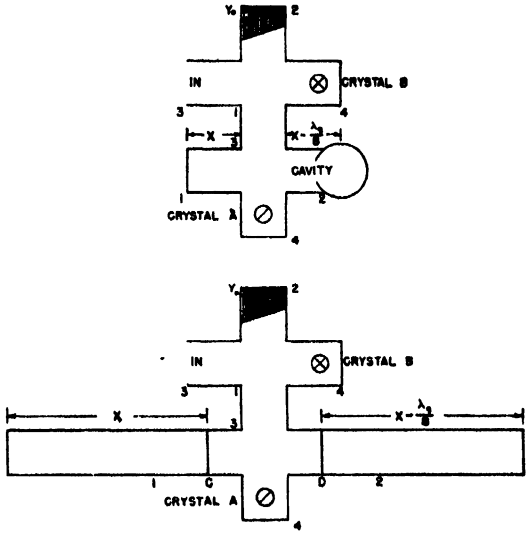


FIG. 1. Spectrum-line discriminator and frequency stabilizer.

---

[3] J. H. Van Vleck, "The relation between absorption and the frequency dependence of refraction," Report 735, M.I.T. Radiation Laboratory.

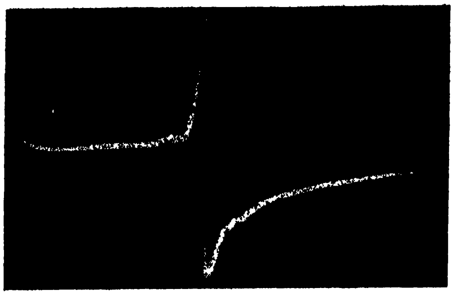


FIG. 2. $NH_3$ 3,3 discriminator characteristic, showing fine structure. The spacing between the center of the linear portion of the discriminator curve and the first satellite is 1.72 megacycles, establishing the frequency scale.

Expanding Eqs. (1), (2), and (3) for $\epsilon_1-1$, $\epsilon_2$, $\Delta\nu/\nu_0$ and $\tanh\gamma l$ all very small compared to 1

$$Y=\frac{1}{Z_2-Z_1}=\frac{2\lambda_g\Delta\nu}{\pi lA\nu_0}\left[\frac{2jQ_0(\nu-\nu_0)}{\nu_0}+1\right]Y_0, \quad (4)$$

where

$$Q_0=\frac{\nu_0}{2\Delta\nu}. \quad \text{If} \quad \alpha=\frac{\pi lA\nu_0}{2\lambda_g\Delta\nu} \quad \text{and} \quad a=\frac{2Q_0(\nu-\nu_0)}{\nu_0},$$

Eq. (4) may be rewritten as

$$Y=\left(\frac{1}{\alpha}+j\frac{a}{\alpha}\right)Y_0, \quad (5)$$

which is Eq. (1) of reference 1 describing the admittance properties of a resonant cavity.

For collision broadening, $A/\Delta\nu$ is independent of pressure.[3] Thus, for a cell of fixed length, as the pressure is lowered the parameter $\alpha$ remains constant while $Q_0$ increases until at sufficiently low pressures effects other than collision broadening begin to influence the line breadth. A discussion of these effects is given by Townes.[4] He states that Doppler effect and collisions with the wave guide walls should begin to affect the line width at $2\Delta\nu=150$ kilocycles, while saturation effects causing an increase in line breadth with input power may limit the line breadth to over one megacycle for high input power.

Townes' data indicate that these saturation effects are just noticeable for an input power of 1.8 milliwatts into an absorption tube of 10.7-mm by 4.3-mm cross section for a pressure of $4\times10^{-2}$ mm of mercury. At this pressure he gets for the ammonia 3,3 line a peak absorption of $5\times10^{-4}$ neper per cm in wave guide and a line width of $2\Delta\nu=1.9$ megacycles. Since the attenuation in nepers/cm $=(2\pi/\lambda_g)\epsilon_2$, the parameters of Eq. (4) are found to be

$$\begin{aligned} Q_0&=\frac{23{,}879}{1.9}=12{,}500, \\ \alpha&=\left(\frac{\pi l\epsilon_2}{\lambda_g}\right)_{\nu=\nu_0}=\frac{l}{3200}. \end{aligned} \quad (6)$$

The slope $dv/d\nu)_{\nu_0}$ of the discriminator curve as derived by Pound is

$$\left.\frac{dv}{d\nu}\right)_{\nu_0}=DP_0\frac{Q_0}{\nu_0}\frac{2\alpha}{(1+\alpha)^2} \quad (7)$$

volts per megacycle, where $P_0$ is the input power to the discriminator and $D$ is a constant. Typical cavity constants are $Q_0=12{,}500$, $\alpha=1$, and $P_0\approx1$ milliwatt, where the upper limit to $P_0$ is set by the amount of decoupling necessary to prevent frequency discontinuities in tying the discriminator to the oscillator. For the spectrum-line discriminator with a 2-meter cell, the constants are $Q_0=12{,}500$, $\alpha=\frac{1}{16}$, and again $P_0=1$ milliwatt, the limit on $P_0$ this time being set by saturation effects. The ratio $dv/d\nu_{\text{spectrum line}}$ to $dv/d\nu_{\text{cavity}}$ can therefore be made at least 1 to 4.5. If Pound's value of $DP=1$ volt is applicable to 1-cm oscillators and crystals as well as to the 3-cm ones to which he applied it, this gives $dv/d\nu=250$ millivolts per megacycle for a cavity and 55 millivolts per megacycle for a spectrum line.

## OBSERVATIONS

Figure 2 is a photograph of the $NH_3$ 3,3 discriminator characteristic at a pressure of about one one-hundredth of a millimeter of mercury. The fine-structure satellites of the line are clearly visible, two on each side of the discriminator. The spacing between the center of the linear portion of the discriminator curve and the first satellite is 1.72 megacycles,[5] establishing the frequency scale. The slope of a similar characteristic measured at a somewhat higher pressure was 70

---

[4] C. H. Townes, Phys. Rev. 70, 665 (1946).

[5] B. P. Dailey, R. L. Kyhl, M. W. P. Strandberg, J. H. Van Vleck, and E. B. Wilson, Jr., Phys. Rev. 70, 984 (1946).

millivolts per megacycle, comparing favorably with the rough theoretical estimate of 55 millivolts per megacycle.

For purposes of comparison of spectrum-line discriminators and cavity discriminators, the above measured slope *dv/dν* is all the data needed. The ultimate purpose of such a discriminator, however, is to actually stabilize an oscillator. To this end, the microwave oscillator which had been frequency swept to obtain the discriminator characteristic was operated *CW*, and the discriminator output fed into a d.c. amplifier similar to that described by Pound. The amplifier output was fed to the reflector electrode only, and the oscillator frequency was observed as a function of the setting of a potentiometer controlling the reflector voltage. Thus the actual voltage on the reflector was the sum of the potentiometer control and the error signal from the amplifier. It was observed that if the pressure of ammonia in arm 2 was a centimeter or more, or if the ammonia were frozen out at liquid air temperatures, the oscillator frequency was a continuous function of the potentiometer control. If, however, the pressure was in the range corresponding to a line width of one-half to two megacycles, approximately, the frequency would "lock in" to the line frequency of 23,870 megacycles over a range of potentiometer settings, jumping discontinuously to a different frequency when the limit of control of the discriminator was exceeded. At approximately optimum conditions as the potentiometer was continuously varied in such a direction as to tend to increase the oscillator frequency, the oscillator tuned continuously up to 23,841 megacycles, jumped to the line frequency of 23,870 megacycles, remaining constant within one megacycle until it again jumped to 23,892 megacycles. Control was thus maintained for variations in reflector voltage equivalent to a frequency spread of just under 51 megacycles. The measured variation of potentiometer voltage over which the frequency remained constant was 14 volts (of a total reflector-cathode voltage of 75 volts).

# The Efficiency of Thermoelectric Generators. I.*

MARIA TELKES
*Department of Metallurgy, Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Received July 28, 1947)

The generation of electrical energy from thermal energy by thermoelectric means cannot be accomplished with thermocouples made of the available alloys generally used for temperature measurements, their efficiency being less than one percent.

A review of the theoretical efficiency calculations shows that higher efficiencies can be attained with thermocouple materials to which the Wiedemann-Franz-Lorenz relation is applicable, when their thermoelectric power is greater than 200 microvolts/°C. Some zinc-antimony alloys with added metals approach the above conditions and have produced an experimental efficiency in excess of five percent, in accordance with the theoretical calculations.

The criteria of higher efficiency are applied to semiconductors, deriving the optimum conditions. Lead sulfide with excess lead was found to be the only suitable material at present, which in combination with the zinc-antimony alloy produces an efficiency of seven percent.

Higher thermoelectric efficiencies can be produced only by developing new materials which can attain the theoretically required high values of thermoelectric power, low heat conductivity, and low specific resistance.

## SYMBOLS

| | |
|---|---|
| $T$ | temperature (°abs.) |
| $T_h$ | hot junction temperature (°abs.) |
| $T_c$ | cold junction temperature (°abs.) |
| $\Delta t = T_h - T_c$ | temperature difference (°C) |
| $E$ | open circuit e.m.f. of thermocouple, Seebeck e.m.f. (volt) |
| $e = \dfrac{E}{T_H - T_C}$ | the average thermoelectric power in the temperature range of $\Delta t = T_h - T_c$ (volt/°C) |
| $\theta = dE/dT$ | the instantaneous thermoelectric power at $T$ temperature (volt/°C) |
| $I$ | current flowing in thermocouple and external load circuit (amp) |
| $R$ | resistance of thermocouple (ohm) |
| $R_e$ | external load resistance, $R_e = R$ (ohm) |
| $W_0$ | output in external circuit (watt) |
| $W_i$ | thermal energy absorbed at the hot junction (watt) |
| $Q_c$ | heat conductance of thermocouple (watt) |
| $\rho'\rho''$ | specific electric resistance of thermocouple materials average between $T_h$ and $T_c$ (ohm cm) |
| $\sigma$ | specific electric conductivity at $T$ temperature (ohm$^{-1}$ cm$^{-1}$) |
| $k'k''$ | specific heat conductivity of thermocouple materials averaged between $T_h$ and $T_c$ (watt cm$^{-1}$ deg.$^{-1}$) |
| $s's''$ | cross-sectional area of thermocouple materials (cm$^2$) |
| $L$ | length of thermocouple materials, both of same length, (cm) |
| $d'd''$ | form factors of thermocouple materials denoting $s'/L$ and $s''/L$ (cm) |
| $\Delta\epsilon$ | energy level difference (electron volt) |
| $k_0$ | Boltzman's constant ($0.8617 \times 10^{-4}$ electron volt/°C) |
| $l_0$ | mean free path (cm) |
| $n_b$ | the number of bound impurity electrons/cm$^3$ |
| $e_0$ | electronic charge |

## HISTORICAL REVIEW

VARIOUS types of thermoelectric generators have been constructed in the past with the aim of producing a gas-heated low voltage direct-current source of relatively small but constant power output. The early efforts have been summarized in book form by F. Peters in 1908.[1] This summary shows that most of the thermoelectric materials then known have been used experimentally in a large variety of empirical designs. Only very few types of thermoelectric generators have reached the production stage. The best-known commercially produced apparatus was Gülcher's gas-heated thermoelectric generator with an over-all efficiency of one-half of one percent.

In more recent times the construction of thermoelectric generators was revived with the aim of producing a gas-heated power supply for radio receivers and for charging storage batteries. Koch's[2] antimony and German silver thermoelectric generator attained an efficiency of 0.5 percent. Wall[3] used iron-constantan thermoelements, attaining an efficiency of less than 0.05 percent. Milne[4] used thermocouples made of nickel and an alloy of copper-silver-selenium in a gas-heated battery charger for radio re-

---

* Publication No. 20, M.I.T. Solar Energy Conversion Research Project.

[1] F. Peters, *Thermoelemente und Thermosäulen*, Halle 1908.

[2] E. Koch, Helios **28**, 557 (1922).

[3] T. F. Wall, The Electr. Rev. 847 (1927).

[4] H. R. Milne, Brit. Patent 510, 410, and 501, 411 Feb. 24 (1939).

ceivers, producing an efficiency less than one percent.

## THE EFFICIENCY PROBLEM

For the purpose of electrical power production an efficiency of one percent is definitely unsatisfactory. It is true that relatively small electrical power generators used at present are comparatively inefficient. Small steam-electric power generators operate with an efficiency of five percent or less. Power plants used on farms for the generation of electricity for household purposes operate with an efficiency of less than ten percent, which compares with the twenty-five percent efficiency rating of large modern electric power generating stations.

What is the cause of the low efficiency of thermoelectric generators? This question has been analyzed repeatedly in the literature. One of the suggested causes was the relatively imperfect heat transfer from the fuel burner to the hot junctions of the thermoelements. It is obvious that only that part of the heat of combustion is effective which is actually transferred to the hot junctions. Considerable heat may be lost in the flue gases and through the structural elements which are necessary to hold the thermoelements.

Peters[1] described numerous empirical designs which attempted to improve the heat transfer conditions, but there are relatively few measurements available concerning this phase of the problem. The over-all efficiency of the thermoelectric generator is best analyzed as the efficiency of the thermoelectric process itself as distinct from the efficiency of the heat transfer to the hot junctions.

The purely thermodynamic treatment of the thermocouple considers the thermocouple as a reversible heat engine (recent summary by W. F. Roeser[5]). The thermodynamic consideration establishes correlations between the Seebeck, the Peltier, and the Thomson effects disregarding the irreversible processes caused by heat conduction, lateral heat losses, and the Joule heat generated in the operating thermoelements. The thermodynamic treatment alone is insufficient to account for the performance of the thermoelectric generator, because the irreversible effects are unavoidable and predominant in the calculation of the efficiency.

The efficiency of the thermoelectric process (eff.) is defined as the ratio of useful electrical power output ($W_0$) delivered to an external load, to the electrical equivalent of the thermal energy absorbed at the hot junctions of the thermoelements ($W_i$)

$$\text{eff.} = W_0/W_i.$$

This definition has been used in previous calculations by Lord Rayleigh,[6] Kollert,[7] and Altenkirch.[8] The knowledge of the following physical properties of the thermoelectric materials is required: the thermoelectric power ($e$), the specific electrical resistance ($\rho$), and the specific heat conductivity ($k$) which are functions of the temperature. The above factors are the material constants of the thermoelements. The form factors (cross section, length, shape factors) may be selected according to certain conditions of best design. The previous calculations failed to produce any correlation between the calculated and the observed efficiency, which was primarily due to the inadequate data concerning the material constants.

In recent years the construction of radiation-measuring thermopiles has been improved considerably. Radiation-measuring thermocouples or thermopiles may be regarded as miniature thermoelectric generators, operating at a relatively small temperature difference. In most cases a high voltage output is desired at the expense of the efficiency. The efficiency calculations for thermoelectric generators and radiation-measuring thermopiles are otherwise identical. Such calculations have been published by Johansen,[9] Firestone,[10] Harris,[11] Cartwright,[12] Jones,[13] Kovalenko,[14] Emmons,[15] and others.

The theoretical calculations mentioned above are revised on the following pages, with the aim of deriving equations and conclusions concerning

---

[5] American Institute of Physics, *Temperature, Its Measurement and Control in Science and Industry* (Reinhold Publishing Corporation, New York, 1941), p. 180.

[6] Lord J. W. Rayleigh, Phil. Mag. **34**, 481 (1892).

[7] J. Kollert, Elektrotech. Zeits. **11**, 333 (1890).

[8] E. Altenkirch, Physik. Zeits. **10**, 560 (1909).

[9] E. S. Johansen, Ann. d. Physik **33**, 517 (1910).

[10] F. A. Firestone, Rev. Sci. Inst. **1**, 630 (1930).

[11] L. Harris, Phys. Rev. **45**, 635 (1934).

[12] C. H. Cartwright, Rev. Sci. Inst. **1**, 592 (1930).

[13] R. V. Jones, J. Sci. Inst. **11**, 247 (1934).

[14] V. Kovalenko, Tech. Phys. U.S.S.R. **5**, 789 (1938).

[15] H. Emmons, J. Frank. Inst. **229**, 29 (1940).

the theoretical efficiency and the best design criteria of thermoelectric generators. During these calculations several assumptions are made which are now considered in detail.

### 1. Thermocouple Materials

The thermoelectric generator consists of a number of thermocouples, which are identical. The theoretical efficiency is not influenced by the number of the thermocouples and therefore in the following calculations only one thermocouple is considered. The couple is formed of two thermoelectric materials, with hot and cold junctions maintained at definite temperatures. The materials are assumed to be in the polycrystalline state. Their specific electric resistance ($\rho'\rho''$), specific heat conductivity ($k', k''$), and thermoelectric power ($e$) are known functions of the temperature ($T$). The materials should be suitably annealed and their physical properties should not change during the operation of the thermocouple

### 2. Junctions

A selection of solders, fluxes, or welding techniques is available for the preparation of satisfactory junctions It is important that the junction should be free of any additional electric resistance, caused by imperfect soldering or welding, due to the presence of oxide layers or other causes. It is assumed that the "junction resistance" is negligible compared to the total resistance of the thermocouple. This condition is relatively easily attained at the cold junction, but difficulties may arise at the hot junctions. During prolonged operation at relatively higher temperatures, thermal corrosion and deterioration may occur at the hot junction, increasing the electrical resistance and thereby decreasing the efficiency.

It is obvious that the hot junction cannot be operated above a certain temperature, limited by the melting points of the thermoelements. Actually the operating temperature of the hot junction is further limited by the nature of the thermoelectric materials and the contact at the hot junction. Even if oxidation is avoided by limiting the hot junction temperature, changes may occur during prolonged operation because of the diffusion of the two thermoelectric materials and the connecting layer. If the hot junctions are formed with the help of a solder, the components of the solder may diffuse into the thermoelectric materials, with the possibility of changing their composition and properties. Such changes may not be noticed except during prolonged life tests The effect of thermal diffusion may be minimized by using welding techniques or by limiting the amount of solder to a minimum, using a soldering material which does not interfere with the physical characteristics of the thermoelectric materials.

### 3. Heating the Hot Junctions

The experimental determination of the efficiency of thermocouples requires a controlled heating arrangement delivering measured amounts of thermal energy to the hot junctions. For experimental purposes electrical heating of the hot junctions is preferable

### 4. Cooling the Cold Junctions

Because of the irreversible heat conduction occurring in the thermoelectric materials a considerable amount of heat is conducted from the hot junction to the cold junction If this energy is not removed by suitable cooling, the temperature of the cold junction will increase The temperature of the cold junctions can be maintained at or near room temperature by using suitable water or air cooling. The use of higher cold junction temperatures is not desirable in thermocouples made of metallic conductors, because their electrical resistance increases with increased temperature, decreasing their efficiency

### 5. Lateral Heat Losses

It is obvious that the lateral or radial heat losses from the surface of the thermoelements should be as small as possible. In a thermoelectric generator, consisting of a large number of thermoelements, these losses may be minimized by arranging the thermocouples in a closely packed assembly, with the unavoidable minimum of insulating material between the couples. Experimental tests made with relatively few or with individual couples may minimize the lateral heat losses, by surrounding the thermocouples with adequate thermal insulation, or the lateral heat losses may be determined experimentally and corrections applied. During the efficiency calculations it is assumed that the lateral heat losses are negligible.

### 6. Form Factors

The thermoelectric materials are generally in the form of cylindrical or rectangular bars, wires, or sheets. Some constructors used wedge-shaped or tubular thermoelements with the aim of improving the performance. Actually, however, such forms failed to increase the efficiency. The construction is simplified if the lengths of the thermoelectric materials are equal, as in this case the distance between the hot and cold junction zones is fixed, facilitating the structural arrangement. Complex shape or form factors alone cannot improve the efficiency.

It can be shown by calculation that optimum conditions will prevail if the shape factor of the thermoelements is selected in such a way that equal amounts of heat are conducted away by the two thermoelectric materials, while at the same time the electric resistance of the two elements is also equal.

### 7. Internal and External Resistance

The maximum power output of a thermoelectric generator is obtained when its internal resistance is nearly equal to the external load resistance This criterion can be derived by calculation and has been proven by experiments.[3]

### 8. Joule Heat

During the operation of a thermoelectric generator with current flowing through the thermoelements, an internal Joule heat will be generated. Part of this heat will be removed at the cold junctions. This effect may be disregarded as long as the efficiency of the thermocouple is rather low.

### 9. Thomson Effect

During the operation of a thermoelectric generator on closed circuit the Thomson effect should influence the thermal distribution conditions. The effect is of relatively slight importance in most cases and therefore it is not considered in the following calculations.

### EFFICIENCY CALCULATIONS

The calculations apply to a definite operating temperature interval $\Delta t = T_h - T_c$.

The open circuit e.m.f. is $E = e\Delta t$. When the circuit is closed through an external resistance ($R_e$) which is equal to the internal resistance ($R$) of the thermocouple a current ($I$) will flow in the circuit: $I = E/2R$. The load e.m.f. will drop to $\frac{1}{2}E$. The Watt output in the external circuit will be:

$$W_0 = \tfrac{1}{2}EI = E^2/4R. \tag{1}$$

The total electrical energy produced in the thermocouple and the external load circuit is $2W_0$. A part of the thermal energy ($W_t$) absorbed at the hot junction will be used for the production of $2W_0$ electrical energy, while the larger part of $W_t$ will be conducted away by the thermocouple elements in accordance with their heat conductivity ($Q_c$) (irreversible loss). Applying the second law of thermodynamics, the amount of thermal energy required for the production of $2W_0$ electrical energy is:

$$\frac{2W_0T_h}{T_h - T_c} = \frac{2W_0T_h}{\Delta t}. \tag{2}$$

The efficiency is obtained by combining Eqs. (1) and (2):

$$\text{eff} = \frac{W_0}{W_t} = \frac{W_0}{\dfrac{T_h 2W_0}{\Delta t} + Q_c} = \frac{1}{\dfrac{2T_h}{\Delta t} + \dfrac{Q_c}{W_0}}$$

$$= \frac{1}{\dfrac{2T_h}{\Delta t} + \dfrac{Q_c 4R}{E^2}}. \tag{3}$$

We proceed substituting the values of the specific resistance, the specific heat conductivity $k'k''$, and the form factors $d'd''$ into the above equation:

The electric resistance

$$R = \rho'/d' + \rho''/d''. \tag{4}$$

The heat conductance

$$Q_c = \Delta t(k'd' + k''d''). \tag{5}$$

TABLE I. Calculated efficiency of thermocouples (percent); cold junction temperature 27°C; Wiedemann-Franz-Lorenz relation applicable.

| Temp. diff. $t$°C | Thermoelectric power in microvolts/°C 100 | 150 | 200 | 250 | 300 | 400 | 500 | 600 | 800 | 1000 |
|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 0.75 | 1.56 | 2.53 | 3.54 | 4.55 | 6.3 | 7.7 | 8.7 | 10.1 | 10.8 |
| 200 | 1.30 | 2.70 | 4.35 | 6.04 | 7.69 | 10.5 | 12.7 | 14.3 | 16.0 | 17.4 |
| 300 | 1.72 | 3.57 | 5.71 | 7.92 | 10.0 | 13.6 | 16 2 | 18.2 | 20.6 | 22.0 |
| 400 | 2.06 | 4.25 | 6.78 | 9.35 | 11 8 | 15.8 | 18.8 | 21.0 | 23 8 | 25 3 |
| 500 | 2.34 | 4.80 | 7.63 | 10.5 | 13.2 | 17.6 | 20.9 | 23.2 | 26.2 | 27.8 |

From Eqs. (4) and (5) it follows that:

$$\frac{4RQ_c}{E^2}=\frac{4\Delta t(k'd'+k''d'')\left(\frac{\rho'}{d'}+\frac{\rho''}{d''}\right)}{e^2\Delta t^2}$$

$$=\frac{4\left(k'\rho'+k''\rho''+\frac{k'\rho''d'}{d''}+\frac{\rho'k''d''}{d'}\right)}{e^2\Delta t}. \qquad (6)$$

Before proceeding further we examine the optimum conditions for the selection of the form factors. It was stated above that the form factors should be selected in such a way that equal amounts of heat should flow through the two elements and that their electrical resistance should be the same. This condition will prevail if:

$k'd'=k''d''$ or $d''/d'=k'/k''$ (from Eq. 5),

$\rho'/d'=\rho''/d''$ or $d''/d'=\rho''/\rho'$ (from Eq. 4),

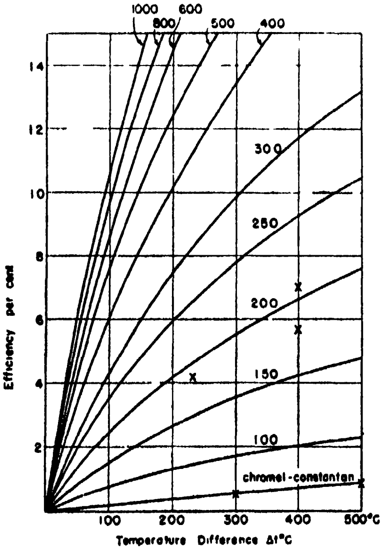


FIG. 1. Calculated efficiency of thermocouples as a function of the temperature difference; x-experimental results. The figures indicate the thermoelectric power in microvolts/°C.

combining the above

$$\frac{d''}{d'}=\left(\frac{k'\rho''}{k''\rho'}\right)^{\frac{1}{2}} \qquad (7)$$

substituting Eq. (7) into (6) we obtain:

$$\frac{4RQ_c}{E^2}=\frac{4\{k''\rho''+k'\rho'+2(k''\rho''k'\rho')^{\frac{1}{2}}\}}{e^2\Delta t}$$

$$=\frac{4\{(k''\rho'')^{\frac{1}{2}}+(k'\rho')^{\frac{1}{2}}\}^2}{e^2\Delta t}. \qquad (8)$$

Combining Eq. (3) and (8) the efficiency is:

$$\text{eff.}=\frac{1}{\frac{2T_h}{\Delta t}+\frac{4\{(k'\rho')^{\frac{1}{2}}+(k''\rho'')^{\frac{1}{2}}\}^2}{e^2\Delta t}}. \qquad (9)$$

Examining the efficiency equation (9) we find the expressions $k'\rho'$ and $k''\rho''$. On the basis of experimental observations, Wiedemann-Franz found (in 1853) that for many metals and alloys the ratio of the heat conductivity ($k$) and the electric conductivity ($\sigma$) is constant at room temperature, $k/\sigma=7.3\times10^{-6}$. As the specific resistance $\rho=1/\sigma$ it equally follows that $k/\sigma=k\rho=7.3\times10^{-6}$.

The specific resistance of many metals and alloys was found to increase with temperature. On the basis of further experimental evidence Lorenz (in 1877) found the general relation (Wiedemann-Franz-Lorenz-constant):

$$k/6=k\rho=2.45\times10^{-8}T. \qquad (10)$$

The theoretical explanation for the value of this constant has been given by Sommerfeld (in 1927).

The experimental evidence shows that the Wiedemann-Franz-Lorenz relation is applicable to most metals and many alloys (within the temperature range of thermocouple use). No metal or alloys have been found to show a lower value. Some alloys show a higher value, a deviation ($D$) from the normal Wiedemann-Franz-Lorenz constant. The deviation ($D$) may be explained by a consideration of the heat conductivity ($k$). The heat conductivity is due to two factors: the atomic part, which is of the same order of magnitude as that of insulators, and the electronic part, which in good heat conductors is preponderant. Metals (Bi, Sb) or alloys of low heat conductivity generally show a deviation ($D$) from the normal Wiedemann-Franz-Lorenz constant.

According to Eq. (9) the efficiency will be high if $k'\rho'$ and $k''\rho''$ are small, that is, if the Wiedemann-Franz-Lorenz constant is applicable to the thermocouple.

Optimum condition: *Wiedemann-Franz-Lorenz relation* applicable. Assuming that Eq. (10) is applicable to a thermocouple operating at an average temperature of $1/2(T_h+T_c)$ it follows that $\rho'k'=\rho''k''=2.45\times10^{-8}1/2(T_h+T_c)$.

In this case Eq. (9) is changed to:

$$\text{eff.}=\frac{1}{\dfrac{2T_h}{\Delta t}+\dfrac{16\rho k}{e^2\Delta t}}=\frac{1}{\dfrac{2T_h}{\Delta t}+\dfrac{19.7\times10^{-8}(T_h+T_c)}{e^2\Delta t}}. \quad (11)$$

This equation contains only the junction temperatures and the thermoelectric power. The cold junction temperature, $T_c$, may be regarded as fixed near the ambient or room temperature, which is about 300°abs. (27°C). Equation (11) permits the calculation of the efficiency for definite temperature differences and thermoelectric powers. The results of this calculation are summarized in Table I and Fig. 1.

It should be emphasized that the above efficiencies represented in Table I and Fig. 1, are based upon the assumption that thermoelectric materials can be produced to which the Wiedemann-Franz-Lorenz relation is applicable. At higher efficiencies (> 10 percent) the Joule heat cannot be neglected and therefore these values should be regarded as tenative indications.

The applicability of the Wiedemann-Franz-Lorenz relation for most metallic materials and alloys can be determined if the specific resistance and specific heat conductivity are known. For a temperature of 27°C (300°abs.) the Wiedemann-Franz ratio $\rho k=7.35\times10^{-6}$. For most materials the specific resistance is known, but the heat conductiyity may not be available. In such cases the following rule may be applied to determine whether a large deviation from the Wiedemann-Franz-Lorenz relation can be expected. The specific heat conductivity may be assumed to be as low as that of certain insulators, which average $k=0.01$. With this assumed value the corresponding $\rho$ value will be $7.35\times10^{-4}$. ***Any material which shows a specific resistance larger than 0.001 ohm-cm at room temperature probably cannot be expected to follow the Wiedemann-Franz relation.***

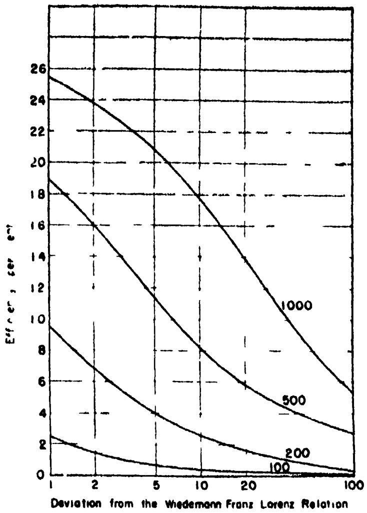


FIG. 2. Calculated efficiency of thermocouples in relation to their deviation from the Wiedemann-Franz-Lorenz relation; temperature difference 400°C; the figures indicate the thermoelectric power in microvolts/°C.

General condition: *Wiedemann-Franz-Lorenz relation not applicable.* Many thermoelectric materials of high thermoelectric power show a considerable deviation ($D$) from the Wiedemann-Franz-Lorenz relation. If the specific resistance, $\rho'$, and the specific heat conductivity, $k'$, of such materials is known, the deviation ($D$) at an average temperature ($T$) will be:

$$D=\frac{\rho'k'}{2.45\times10^{-8}T}.$$

A thermocouple formed of a thermoelectric material showing a deviation ($D$) in combination

TABLE II. Calculated efficiency of thermocouples (percent); cold junction temperature 27°C, temperature difference 400°C; $D$ = Deviation from the Wiedemann-Franz-Lorenz relation.

| Thermoelectric power microvolt/°C | $D=1$ | 2 | 5 | 10 | 20 | 100 |
|---|---|---|---|---|---|---|
| 100 | 2.1 | 1.3 | 0.8 | 0.5 | 0.3 | 0.07 |
| 250 | 9.4 | 6.7 | 4.2 | 2.7 | 1.6 | 0.4 |
| 500 | 18.8 | 15.8 | 11.6 | 8.3 | 5.5 | 1.6 |
| 1000 | 25.3 | 23.7 | 21.0 | 17.8 | 14.0 | 5.4 |

Table III. Commercially available temperature indicating thermocouples.

| | Thermoelectric power ($e$) microvolts/°C Temp range 27–727°C |
|---|---|
| Platinum alloys | less than 20 |
| "Pallaplat" couple[18] | 55 |
| Chromel–alumel | 42 |
| Iron-constantan | 57 |
| Chromel P constantan | 76 |

with another material showing no deviation will show a diminished efficiency, as computed in Table II and shown in Fig. 2. The calculations were made for a cold junction temperature at 27°C, using a temperature difference of 400°C.

Table II indicates the importance of the deviation from the Wiedemann-Franz-Lorenz relation, which is very pronounced for the 100- to 250-microvolt/°C thermoelectric power values; the efficiency of such couples diminishes to half of the theoretical maximum with a deviation of two to five. For the 500- to 1000-microvolt/°C range the efficiency decreases to half of the theoretical value with a 10- to 20-fold deviation. Assuming that such materials would have a low specific heat conductivity of $k=0.01$, the corresponding $\rho$ value will be about 0.03 ohm-cm. *It is probable that in the above indicated temperature range thermoelectric materials with a specific resistance greater than 0.03 ohm-cm could not attain an efficiency in excess of 10 percent, even if their thermoelectric power were as high as 1000-microvolt/°C.*

## THERMOCOUPLE MATERIALS

Recent reviews[6,16,17] tabulate the properties of thermoelectric materials used for temperature measurements. Some of the best materials of this type are summarized in Table III.

As shown in Table III, the thermoelectric power of the commercially available thermocouples is considerably below 100-microvolt/°C. Even if the Wiedemann-Franz-Lorentz relation of these materials would not deviate from its normal value, the expected efficiency could not be greater than 2 percent, operating them at relatively high temperatures.

### The Chromel P–Constantan Thermocouple

This couple produces the highest thermoelectric power and may be regarded as the best commercially available couple. It was used, therefore, to verify the efficiency calculations, comparing them with the experimentally attainable efficiency. Chromel P (approximately 90 nickel–10 chromium) and constantan (55 copper–45 nickel) were used in the form of flat bars with welded hot and cold junctions. The hot junctions were electrically heated, the cold junctions water-cooled. The specific resistance, heat conductivity, and thermoelectric power of these alloys are shown in Table IV together with their calculated and experimentally observed efficiencies.

The calculated and observed efficiencies are in good agreement and show that the efficiency is less than one percent with the hot junctions operating at 520°C. This is approximately the highest temperature at which constantan can be used continuously without undue thermal corrosion. Assuming that constantan could be protected against oxidation at higher temperatures, the calculations show that an efficiency of two percent could be attained with a temperature difference of about 1000°C.

### The Bismuth-Antimony Couple

The use of this couple is limited by the melting point of bismuth (271°C). Instead of bismuth, an

Table IV.

| | | Temperature range | |
|---|---|---|---|
| | | 20–320°C | 20–530°C |
| Constantan* | | | |
| specific resistance | $\rho$ | $49\times10^{-6}$ | $49\times10^{-6}$ |
| heat conductivity | $k$ | 0.29 | 0.33 |
| | $\rho k$ | $14.3\times10^{-6}$ | $16.2\times10^{-6}$ |
| deviation | $D$ | 1.30 | 1.20 |
| Chromel P | | | |
| specific resistance | $\rho$ | $75\times10^{-6}$ | $78\times10^{-6}$ |
| heat conductivity | $k$ | 0.19 | 0.22 |
| | $\rho k$ | $14.3\times10^{-6}$ | $17.2\times10^{-6}$ |
| deviation | $D$ | 1.30 | 1.28 |
| Chromel P–constantan | | | |
| thermoelectric power (volt/°C) | | $71\times10^{-6}$ | $73\times10^{-6}$ |
| calculated efficiency (percent) | | 0.64 | 0.96 |
| observed efficiency (percent) | | 0.58 | 0.92 |

* Constantan and alloys of nearly identical composition are manufactured under various trade names. Advance, Ideal, Copel.

---

[16] A. W. Smith, Ohio State Univ. Bull. No. 20. Eng. Exp. Station (1921).

[17] G. Borelius, Handbuch der Metallphysik 1, 185 (1935).

[18] A. Schulze, *Metallische Werkstoffe für Thermoelemente* (N. E. M.-Verlag Dr. G. Lüttke, Berlin, 1940).

alloy of 91 Bi–9 Sb can be used which shows a slightly better performance. The electrical properties of these materials have been determined by Gehlhoff and Neumeier[19,20] and are summarized in Table V, giving the calculated efficiency.

## THERMOELECTRIC PROPERTIES OF ELEMENTS AND ALLOYS

The thermoelectric power and the specific resistance of various alloys have been summarized.[5,16,17,19–23]

According to the literature the thermoelectric power of most elements is considerably less than 100-microvolt/°C, with the exception of Te, Se, Si, and Ge.

The thermoelectric power of most alloys was found to be considerably less than 100-microvolt/°C, therefore these alloys could not be expected to produce efficiencies higher than two to three percent, even when operated at relatively high temperature differences.

Thermoelectric power values in excess of 100-microvolt/°C apparatently are found only in alloys containing Sb, Bi, Se, Te, Si (group IVB, VB, VIB of the periodic table), in combination with other metals.

TABLE V.

| | | Temperature range 260–20°C |
|---|---|---|
| Bismuth | | |
| specific resistance | $\rho$ | $185\times10^{-6}$ |
| heat conductivity | $k$ | 0.096 |
| | $\rho k$ | $17.8\times10^{-6}$ |
| deviation | $D$ | 1.74 |
| Antimony | | |
| specific resistance | $\rho$ | $86\times10^{-6}$ |
| heat conductivity | $k$ | 0.205 |
| | $\rho k$ | $17.8\times10^{-6}$ |
| deviation | $D$ | 1.74 |
| 91 Bi–9 Sb | | |
| specific resistance | $\rho$ | $212\times10^{-6}$ |
| heat conductivity | $k$ | 0.073 |
| | $\rho k$ | $15.4\times10^{-6}$ |
| deviation | $D$ | 1.51 |
| Bismuth-antimony couple | | |
| thermoelectric power (volt/°C) | | $111\times10^{-6}$ |
| calculated efficiency (percent) | | 1.00 |
| 91 Bi–9 Sb *vs.* antimony couple | | |
| thermoelectric power (volt)$e$ | | $123\times10^{-6}$ |
| calculated efficiency (percent) | | 1.31 |

Operating at the highest attainable temperature difference, these thermocouples show a calculated efficiency at 1.00 and 1.31 percent.

TABLE VI. Properties of tellurium and tellurium alloys.

| Composition weight percent | Spec. resist. $\rho$ (at room temp.) | Heat cond. $k$ (at room temp.) | Thermoelectric power microvolt/°C ref. Cu 100–20°C temp. range | References (see Bibliography) |
|---|---|---|---|---|
| Te (pure) | 0.04 | 0.014 | 70 | 24 |
| Te (pure) | 0.025 | | 70 | 26 |
| Te (impure) | 0.035 | 0.018 | 300 to 500 | 24 25 |
| Te (impure) | 0.009 to 0.40 | | 300 to 550 | 25 |
| 1 Bi bal. pure Te | 0.0024 | | 220 | 26 |
| 0.5 Sb bal. pure Te | 0.0032 | | 215 | 26 |
| $Bi_2Te_3$(52Bi–48Te) | | | | |
| impure | 0.0020 | | 120 | 27 |
| impure | 0.0010 | | 170 | 26 |
| pure | 0.0008 | | −150 | 26 |
| 42Bi–58Te pure | 0.0009 | | −160 | 26 |
| 42Bi–58Te impure | 0.00084 | | −100 | 27 |

A thermocouple formed of 1Bi–99Te and 42Bi–58Te shows a thermoelectric power of 380 microvolt/°C. Efficiency experiments are in progress.

The following general rules have been derived from the study of binary alloy systems:

In a binary alloy series a high value of the thermoelectric power and a high value of the specific electric resistance generally occurs at or near an intermetallic compound.

A relatively small amount of added metal ("impurity") may diminish the specific resistance of an element or an intermetallic compound to a considerable extent, while the thermoelectric power is changed relatively slightly, the heat conductivity remaining practically unchanged.

### Tellurium and Tellurium Alloys

The general rules indicated above can be illustrated by tellurium and its alloys. The thermoelectric power of pure tellurium is rather low, while its specific resistance is high. Impure tellurium generally shows a high thermoelectric power, with variable specific resistance. Certain small additions (Bi, Sb) diminish the specific resistance about 100-fold, increasing the thermoelectric power of pure tellurium. The intermetallic compound $Bi_2Te_3$(52 Bi–48 Te) when

---

[19] G. Gehlhoff and F. Neumeier, Verh. d. D. Phys. Ges. 15, 1069 (1913).

[20] G. Gehlhoff and F. Neumeier, Verh. d. D. Phys. Ges. 15, 876 (1913).

[21] W. Guertler, Zeits. f. Metallkunde 22, 78 (1930).

[22] W. Meissner, Thermoelektrizität in Vol. 11, part 2, in Wien-Harms *Handbuch der Experimentalphysik* (Akademische Verlagsgesellschaft, Leipzig 1935).

[23] M. A. Hunter, p. 1227 of reference 5.

[24] C. H. Cartwright, Ann. d. Physik 18, 656 (1933).

[25] C. H. Cartwright, Proc. Roy. Soc. London A148, 648 (1935).

[26] M. Telkes, unpublished.

[27] W. Haken, Ann. d. Physik 32, 291 (1910).

TABLE VII. Properties and efficiency of Zn–Sb alloys; temperature interval 20 to 420°C.

|  | Spec resist ($\rho$) | Spec. heat conductivity ($k$) | Thermoelectric power ref. constantan ($e$) | Efficiency (percent) |  |
|---|---|---|---|---|---|
|  |  |  |  | observed | calculated |
| ZnSb | 0.0072 | 0.014 | 250 | 2.7 | 2.7 |
| ZnSb with small amounts of Sn, Ag | 0.0027 | 0.015 | 250 | 5.6 | 5.7 |
| ZnSb with other additions | 0.0022 | 0.015 | 220 | 4.8 | 4.8 |
|  | 0.0044 | 0.015 | 280 | 5.2 | 5.1 |

pure has a negative thermoelectric power, which changes sign when impurities are present. Another alloy 42 Bi–58 Te shows the highest negative thermoelectric power of this series. The properties of some Te alloys are shown in Table VI.

### Zinc-Antimony Alloys

It has been known[1] that an alloy of Zn and Sb forming the intermetallic compound ZnSb (35 Zn–65 Sb) is a suitable material for thermoelements. The specific resistance, specific heat conductivity, and thermoelectric power as a function of the temperature has been measured.[28, 29] Relatively small amounts of metals added to the ZnSb compound diminish its specific resistance considerably, while the thermoelectric power and the specific heat conductivity are altered only slightly. Efficiency measurements have been made with thermocouples of ZnSb with constantan, at a temperature difference of 400°C, the maximum allowable operating range. The observed efficiency values agree with the calculated values (Table VII).

A thermocouple formed of the above alloys in combination with an alloy of 91 Bi+9 Sb, Table V, attained an efficiency of 4.2 percent operating at a temperature difference of 230°C (hot junction at 250°C) which is the highest allowable temperature due to the low melting point of the bismuth alloy (near 270°C). At this relatively low temperature thermal corrosion is entirely absent. The Zn–Sb alloy thermocouples in combination with constantan or with the Bi–Sb alloy show a considerable improvement in efficiency (4.2 to 5.6 percent) compared with Chromel P–constantan and the Bi–Sb couples (about 1 percent).

### Cadmium-Antimony Alloys

The Cd–Sb series of alloys form an unstable intermetallic compound. This condition influences greatly the electrical properties, as reported in the literature.[30, 31] Efficiency tests failed to produce favorable results.

### Silicon and Silicon Alloys

The thermoelectric power and the resistivity of silicon is very markedly influenced by impurities. Relatively small changes in the amount of the impurities may result in a reversal of the sign of the thermoelectric power.[32–34] The specific heat conductivity of impure silicon is comparable to that of nickel, while the specific resistance of silicon is at least 1000 times greater. The Wiedemann-Franz-Lorenz relation, therefore, is very unfavorable and a high efficiency could not be expected.

### Sulfides, Oxides, Semi-Conductors

The sulfides and oxides as thermoelectric materials may be classified according to the nature of their electrical conductivity.

*Metallic conductors* of high conductivity, which decreases at increasing temperatures. Their temperature coefficient of resistivity is negative, as it is in most metals and alloys.

*Semi-conductors* of comparatively low electrical conductivity, which increases considerably at higher temperatures; the temperature coefficient of resistivity is negative. The thermoelectric power of semi-conductors is generally in the range of 100- to 1000-microvolt/°C.

A large group of semi-conductors show pure electronic conductivity, and only these are considered. Their electrical properties are usually

---

[28] M. Telkes, U. S. patent 2,229,482 (1941).

[29] M. Telkes, U. S. patent 2,366,881 (1945).

[30] A. Eucken and G. Gehlhoff, Ber. d. D. Phys. Ges. **14**, 169 (1912).

[31] F. Fischer and G. Pfleiderer, Ges. Abh. Kennt. Kohle **4**, 409, 440 (1919).

[32] Broniewski and Smialowski—Rev. Met. **29**, 542, 602 (1932).

[33] M. Schulze, Zeits. f. Metallk. **23**, 261 (1931); Physik. Zeits. **31**, 1062 (1930); Zeits. f. tech. Physik **11**, 443 (1930).

[34] F. Fischer and Dehn, Ges. Abh. Kenntn. Kohle **12**, 526 (1937).

influenced by the presence of small amounts of impurities. With the exception of the elements Te, Se, Si, and Ge the semi-conductors are chemical compounds or intermetallic compounds. In their pure and stochiometrically exact form their electrical conductivity shows a minimum, approaching that of insulators.

Semi-conductors of the oxide and sulfide group are usually prepared by compressing and sintering. This procedure generally does not furnish solid compacts, a certain porosity remains, which decreases the conductivity, although it does not influence the thermoelectric power. With suitable treatment the semi-conductors can be prepared in a non-stochiometric form, with the electropositive or the electronegative element in excess as impurity (homogeneous or intrinsic impurity). Added foreign atoms in a similar way can act as heterogenous (extrinsic) impurities. The impurities may supply free conduction electrons or may combine with free electrons, altering the conductivity of the pure stochiometric compound. In some cases the conductivity may increase $10^3$- to $10^6$-fold. The thermoelectric power is generally not affected in a similar manner; in some cases it is hardly influenced at all.

The semi-conductors are classified as:

a. excess electron conductors containing a stochiometric excess of a metallic element; they generally show a negative thermoelectric power. The excess metal may be obtained through a reducing heat treatment.

b. defect or hole conductors containing a stochiometric excess of a non-metallic element; they generally show a positive thermoelectric power. The excess non-metal may be obtained by heat treatment in a suitable gas.

The theoretical treatment of the electrical conductivity and the thermoelectric power of semi-conductors[35–40] results in the following correlation:

The electrical conductivity

$$\sigma = A\exp(-\Delta\epsilon/2k_0T),$$

where

$$A = 0.024 l_0 n_b^{\frac{1}{2}} T^{\frac{3}{4}} (\text{ohm}^{-1}\,\text{cm}^{-1}).$$

The thermoelectric power

$$\theta = \pm[(\Delta\epsilon/2Te_0)+\text{const.}].$$

The value of the constant is $5k/4e_0 = 108\times10^{-6}$ volt for extrinsic conductors and $2k/e_0 = 172\times10^{-6}$ volt for intrinsic conductors.

The above equations may be verified experimentally by measuring the electrical conductivity ($\sigma$) as a function of the temperature ($1/T$) and deriving the values of $A$ and $\Delta\epsilon$. The experimental determination of $\theta$ as a function of the temperature ($1/T$) should result in the same $\Delta\epsilon$ value.

Several attempts have been made to correlate the theoretical calculations with the experimental evidence. Hochberg[41] found a reasonable agreement in the case of $V_2O_5$ and $WO_3$, but only a qualitative agreement with $MoS_2$, CuO and a mixture of cobalt oxides. Borelius[42, 43] found satisfactory agreement for Se.

The equations for $\sigma$ and $\theta$ are now examined from the point of view of the efficiency of the thermoelectric process. The requirement for a high thermoelectric power $\theta$ implies that $\Delta\epsilon$ must

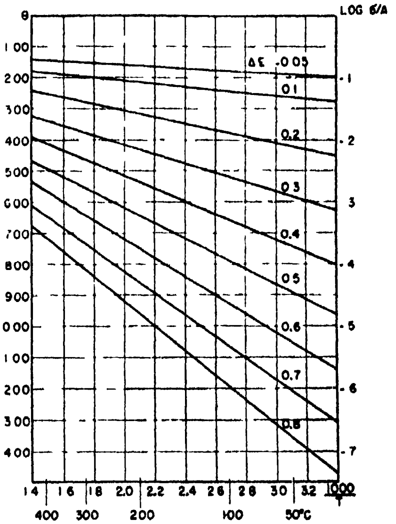


Fig. 3. Thermoelectric power $\theta$ and electric conductivity $\sigma$ of semi-conductors as a function of the temperature $T$ and the energy-level difference $\Delta\epsilon$; $\theta$ in microvolt/°C.

---

[35] A. H. Wilson, Proc. Roy. Soc. **133**, 458 (1931); **134**, 277 (1931).

[36] A. H. Wilson, Semi-Conductors and Metals (Cambridge 1939).

[37] B. Gudden, Ergeb. d. exakt. Naturwiss. **13** (1934).

[38] F. Seitz, J. App. Phys. **16**, 553 (1945).

[39] R. J. Maurer, J. App. Phys. **16**, 563 (1945).

[40] C. Wagner and E. Koch, Zeit. f. physik. Chemie **B32**, 439 (1936).

[41] B. M. Hochberg and M. S. Sominski, Physik. Zeits. Sowjetunion **13**, 198 (1938).

[42] G. Borelius, F. Pihlstrand, J. Andersson, and K. Gullberg, Ark. Mat. Astr. Fys. Svenska Vetens. **2**, No. 14, 30 (1944).

[43] G. Borelius and K. Gullberg, Ark. Mat. Astr. Fys. Svenska Vetens. **2**, No. 17, 31 (1945).

TABLE VIII.

| Material | Electric cond room temp. ohm$^{-1}$ cm$^{-1}$ $\sigma$ | $A$ | $\Delta\epsilon$ | Thermo-electric power $\theta$ microv/°C ref. Cu near room temp. | Author |
|---|---|---|---|---|---|
| $V_2O_5$ monocrystal | $10^{-5}$ | | 0.9 | $-900$ | 41 |
| $V_2O_5$ reduced (?) | 0.03 | | | $-800$ | 48 |
| $WO_3$ sintered | $10^{-6}$ | | 0.9 | $-900$ | 41 |
| $TiO_2$ sintered | 0.2 | 9 | 0.06 | | 45 |
| $TiO_2$ reduced | 5 | 22 | 0.08 | | 44 |
| $TiO_2$ further reduced | 130 | 150 | 0.008 | | 45 |
| $Fe_2O_3$ reduced | 210 | 25 | 0.01 | | 45 |
| $SnO_2$ reduced | 6 | | 0.1 | $-200$ | 46 |
| ZnO sintered reduced | 5 | 5 | 0.03 | $-400$ | 47 |
| PbS pure | $10^{-4}$ | | | 700 | 49 |
| PbS excess Pb | 76 | | | $-300$ | 49 |
| PbS excess S | 8 | | | 100 | 49 |
| PbS melted | 1000 | | | $-104$ | 50 |
| PbS reduced | 455 | | | $-148$ | 50 |
| PbS excess S | 30 | | | 71 | 50 |
| PbS almost pure | 0.3 | | | 530 | 26 |
| PbS excess Pb | 800 | | | $-150$ | 26 |
| PbS excess Pb (max.) | 2000 | | | $-100$ | 26 |
| Ti sulfides sintered | 0.15 | | | $-500$ | 26 |
| Ti excess S | 100 | | | $-160$ | 26 |
| $Cu_2S$ fused or sintered | 20 | | | 280 | 26 |
| $Cu_2S$ excess Cu | 0.03 | | | 620 | 26 |
| $Cu_2S$ excess S | 200 | | | 80 | 26 |

be high. With increasing temperature the thermoelectric power decreases, therefore higher temperatures are less advantageous. The electrical conductivity is subject to the opposite condition, it increases considerably at higher temperatures and decreases at higher values of $\Delta\epsilon$. The value of $A$ should be large, as the electrical conductivity is directly proportional to it. It should be noted that the porosity of pressed and sintered compacts should alter the value of $A$, but should not influence $\Delta\epsilon$.

Using various values of $\Delta\epsilon$, the conductivity in the form of $\log\sigma/A$ and the thermoelectric power $\theta$ can be plotted as a function of the temperature $1/T$. This is shown in Fig. 3. If the values of $\Delta\epsilon$ and $A$ are known for any semi-conductor, its conductivity as a function of the temperature can be obtained from Fig. 3, by inserting $A$ into the corresponding $\sigma/A$ values found on the particular $\Delta\epsilon$ line.

The efficiency calculations indicated that higher efficiencies can be expected only with thermoelectric powers in excess of 300-microvolt/°C, with a low heat conductivity of the order of magnitude $k=0.01$ with low specific resistance of the order of $\rho=0.001$ to $0.03$ (conductivity $\sigma=1000$ to $33$).

Figure 3 shows that near room temperature $\Delta\epsilon=0.2$ would correspond to a thermoelectric power of 440-microvolt/°C, with a corresponding $\log\sigma/A=-1.7$ or $\sigma/A\cong 0.02$. For higher efficiency $\sigma$ should be 33 to 1000 and therefore $A$ should be at least 1500 or greater. At a temperature of 200°C $\Delta\epsilon=0.2$ would correspond to a thermoelectric power of 310-microvolt/°C, with $\sigma/A\cong 0.1$ requiring an $A$ value$>1000$.

The application of the efficiency calculations indicates that higher efficiencies may be obtained only from those semi-conductors, which show $\Delta\epsilon$ of the order of 0.2, with corresponding $A>1000$.

Table VIII shows the properties of some oxides and sulfides with relatively higher thermoelectric power values. The thermoelectric power of some other oxides or sulfides is rather high, but as their conductivity is excessive they were not included in this table.

Mixtures of oxides have been investigated by F. Fischer.[51, 52] The electrical conductivity of certain mixtures is much higher than the conductivity of the components; the thermoelectric power of the highly conducting mixtures is somewhat lower than that of the components.

## Lead Sulfide Thermoelement

Inspecting Table VIII the material of highest conductivity is PbS with excess lead. This mixture shows metallic conduction, the conductivity decreasing at increasing temperatures. PbS thermoelements containing a slight excess of lead may be prepared by hot pressing and sintering. A thermocouple made of PbS and the ZnSb alloys of Table VII shows a thermoelectric power of 330 to 390 microvolt/°C. The specific heat conductivity of PbS was found to be

---

[44] W. Meyer, Physik. Zeits. 36, 749 (1935).

[45] W. Meyer and H. Nedel, Physik. Zeits. 38, 1014 (1937).

[46] G. Bauer, Ann. d. Physik 30, 433 (1937).

[47] O. Fritsch, Ann. d. Physik 22, 375 (1935).

[48] I. K. Krch, Abh Staats Univ. Saratov 14, 54 (1938).

[49] H. Hintenberge, Zeits. f. Physik 119, 1 (1942).

[50] L. Eisenmann, Ann. d. Physik 38, 121 (1940).

[51] F. Fischer *et al.*, Ann. d. Physik 15, 109 (1932).

[52] F. Fischer *et al.*, Ges. Abh. Kenntn. Kohle 12, 516 (1937).

$k = 0.015$. The calculated efficiency of this couple, operating at a temperature difference of $\Delta t = 400°C$, is approximately 10 percent. The experimentally observed efficiency was 7 percent because of difficulties in attaching suitable end contacts to the PbS, which should be free of any contact resistance.

Efficiency tests made with $Cu_2S$ failed to produce any noteworthy results.

## CONCLUSIONS

The efficiency calculations establish a definite relationship between the thermoelectric power ($e$ or $\theta$), the specific electric resistance ($\rho$) or conductivity ($\sigma$), and the specific heat conductivity.

Higher efficiencies may be expected only with materials which show a constant Wiedemann-Franz-Lorenz relationship or a relatively small deviation from it. This criterion requires that the average specific resistance ($\rho$) should not be greater than 0.002 ohm cm ($\sigma = 500$) for materials of medium thermoelectric power (around 200-microvolt/°C). For high thermoelectric power values (around 500- to 1000-microvolt/°C, the maximum permissible average specific resistance ($\rho$) is about 0.03 ohm cm ($\sigma = 33$).

The thermoelectric power of most normal metals and alloys is less than 100-microvolt/°C, therefore their calculated efficiencies cannot exceed two to three percent even when operating at a temperature difference of 500 to 900°C. The observed efficiency of the best commercially available thermocouple (Chromel P-constantan) was only 0.9 percent, operating at the highest allowable temperature difference (500°C).

In the medium thermoelectric power range (200- to 300-microvolt/°C) some of the recently developed zinc-antimony alloys (containing small amounts of added metals) forming a thermocouple with constantan, produced an efficiency of 4.8 to 5.7 percent, at the highest allowable temperature difference (400°C). The same zinc-antimony alloy forming a thermocouple with a bismuth-antimony alloy, produced an efficiency of 4.2 percent at a much lower temperature difference (230°C). For these thermocouples the calculated and observed efficiencies showed very good agreement.

A survey of various alloys failed to produce any thermoelectric materials of higher calculated efficiency, with the exception of the tellurium-bismuth alloys and tellurium, with small amounts of some other metals.

Materials of higher thermoelectric power were found to contain Si, Sb, Bi, O, S, Se, Te (groups IV*B*, V*B*, VI*B* of the periodic table). Some of these materials are semi-conductors. The theoretical correlation of the electric conductivity ($\sigma$) and the thermoelectric power ($\theta$) were applied to derive the criteria of high efficiency for this group of materials. A survey of published data failed to reveal suitable materials of high efficiency, with the exception of PbS containing excess lead, which in this form is actually a metallic conductor. Thermocouples made with PbS (with excess lead) and some of the zinc-antimony alloys mentioned above, produced an efficiency of 7 percent, with a possible expectation of 10 percent indicated by the calculation. The discrepancy is due to a junction-resistance.

Some of the semi-conductor material showing high thermoelectric powers are greatly affected by relatively small amounts of impurities. The presence of impurities may increase the conductivity very considerably, while the thermoelectric power is changed only slightly, the heat conductivity remaining practically uninfluenced. Further study of the effects of impurities may improve the properties of some materials showing high thermoelectric power values.

The theoretical calculations indicate that the problem of producing thermoelectric generators of higher efficiency can be solved only by continued search for materials of high thermoelectric power, low specific electric resistance (high electric conductivity) and low specific heat conductivity.

# Cesium Chromate Photo-Tube Pellets

H. A. Liebhafsky and A. F. Winslow
*Research Laboratory, General Electric Company, Schenectady, New York*

(Received September 10, 1947)

The following results have been obtained in an exploratory investigation of the cesium chromate pellets used in the manufacture of photo-tubes: (1) hydrogen evolution or titration can be used to measure the yield of cesium with reasonable precision. Each method has limitations. (2) The theoretical yield could be obtained from cesium chromate-silicon pellets. (3) Within limits, the actual yield increases with the rate of heating; the maximum yield is not highly sensitive to composition in the range investigated. (4) No marked advantage was observed when titanium or zirconium was substituted for silicon as reducing agent. (5) An excess of hydrogen above the theoretical was consistently observed with nickel pellet holders, which points to a reaction of water vapor with reactive nickel formed during the heating and flashing of the pellet. The substitution of the less volatile molybdenum for nickel eliminated this difficulty. An excess of hydrogen was formed also when the reducing agent in the pellets was titanium.

It was suggested for purposes of further study that the complex process occurring when such pellets are induction heated be subdivided into two reactions of elementary silicon: (1) a preliminary exothermic reduction of hexavalent chromium, which initiates (2) a rapid reduction of monovalent cesium. The two reactions may proceed concurrently. This picture explains several observations made in the laboratory or in the factory.

THE function of the photo-tube pellet is to supply the alkali metal required for the light-sensitive surface. Such a pellet is usually made by compressing an intimately mixed powder containing a reducing agent and a compound of the metal, the metal being liberated in the tube by induction heating. Cesium chromate is a good compound for this purpose, and silicon is a good reducing agent; but little is known about the reaction between them.

Manufacturing processes could be considerably simplified if the pellets gave a reproducible yield. At the present time the only way of providing for enough metal is to make sure of an excess, though an excess is objectionable and must be removed in a subsequent bake-out. This investigation was undertaken in order (1) to find a reliable method of measuring the alkali metal produced when a pellet is "flashed," (2) discover under what conditions cesium chromate-silicon pellets give their maximum yield, and what this maximum is, (3) see how other reducing agents compared with silicon in effectiveness. The results given below were obtained before the war forced the work to be suspended.

## EXPERIMENTAL DETAILS

### Materials

Purified cesium chloride from the laboratory stock, made from pollucite ore by Dr. R. W. Moore, was the starting material for the preparation of cesium chromate. The sparingly soluble dichromate was precipitated out of a chloride solution and redissolved in cesium hydroxide electrolytically prepared, acetone finally being added to obtain the finely divided chromate. Chloride, which is objectionable in a chromate photo-tube pellet, was virtually absent from the product, the principal impurity present being sodium in small amount.

The silicon, titanium, and zirconium were commercial materials of good quality. The titanium was a "hydride;" that is, a product of indefinite composition which liberates hydrogen when heated to a low temperature (*ca.* 200°C) in vacuum. The hydrogen thus liberated did not falsify the experimental results since the system was always pumped out very carefully after the flashing of all pellets (see below).

Since all materials were powders, the particle size is an important variable. The silicon, titanium, and cesium chromate passed a 325-mesh sieve. The zirconium passed 100-, but was caught on 200-mesh. The thoroughly mixed pellet mixture containing silicon or titanium was put through 200-mesh; the zirconium mixture passed 100-mesh.

All pellets were prepared by subjecting about 25 mg of mixture to high pressure in a die. Before being used each pellet was weighed to 0.1 mg on an analytical balance.

## Apparatus

The apparatus in Fig. 1 is designed to produce a mirror of alkali metal in high vacuum when a pellet is heated at a pre-determined rate, to make possible the reaction of this mirror with water vapor, and to permit measurement of the hydrogen thus liberated and titration of the alkali hydroxide formed. The operation of the apparatus is obvious and need not be described in detail. The weight of alkali metal and the pressure of hydrogen will be of the order of 10 mg and 100 microns, respectively; all manipulations were planned to give a precision near 1 percent of these quantities.

The rate of heating the pellet determines the "heating interval," which is the time, measured on a stop watch, that elapses between the closing of the induction-heating circuit and the explosive liberation of metal from the pellet to form a mirror on the enclosing bulb. (The metal escapes around the periphery of the pellet holder, which is held together by 3 or 4 spot welds.) The relationship between the setting of the induction heater and the heating interval had to be established in advance of the actual investigation; since this relationship, which will differ for each experimental arrangement, is sensitive to the relative position of induction coil and pellet, care was taken that this position did not vary from one run to the next. This position was so chosen that the periphery of the pellet holder was heated uniformly in the field. Consistent heating intervals were obtained only when the pellet was held in close contact with the metal, which ensured satisfactory heat transfer.

Pellet holders made of nickel had to be discarded because yields greater than 100 percent cesium were consistently obtained from silicon pellets by the hydrogen method (see below) when they were used; presumably the action of water vapor on active nickel formed during the heating period was responsible. This difficulty disappeared when molybdenum was substituted for nickel.

## Evaluation of Yields

Cesium yields were always calculated according to the gas laws from the hydrogen liberated. To corroborate results thus obtained, the cesium hydroxide was occasionally titrated as follows. 5 cc of standardized 0.02 *N* perchloric acid was pipetted into the reaction bulb *H* (Fig. 1) at the end of the run, and the sides of the bulb were washed down with a little distilled water. After carbon dioxide had been expelled from the acid solution by boiling, the titration was carried out with 0.02 *N* sodium hydroxide containing methyl red. The acid was added to the bulb as soon as possible in order to minimize attack of the glass by cesium hydroxide.

Results from the hydrogen and from the titration method were calculated as percent yield for the pellet, whose weight and initial composition were known. A summary of all pairs of results obtained is given by Fig. 2, those in which the heating intervals were below 4 sec. being distinguished from the others. Inspection of Fig. 2 shows (1) that, three erratic points excepted, the agreement with the 45° line is generally satisfactory, (2) that for the low heating intervals the yield by the hydrogen method tends to exceed that obtained by titration. This tendency probably has the following cause. As the heating interval is decreased, the liberation of cesium becomes more and more explosive, with the result that some metal escapes the reaction bulb.

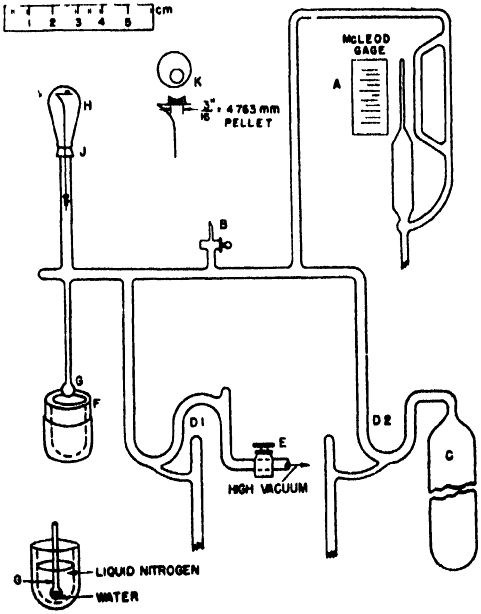


Fig. 1. Apparatus for studying activity of photo-tube pellets.

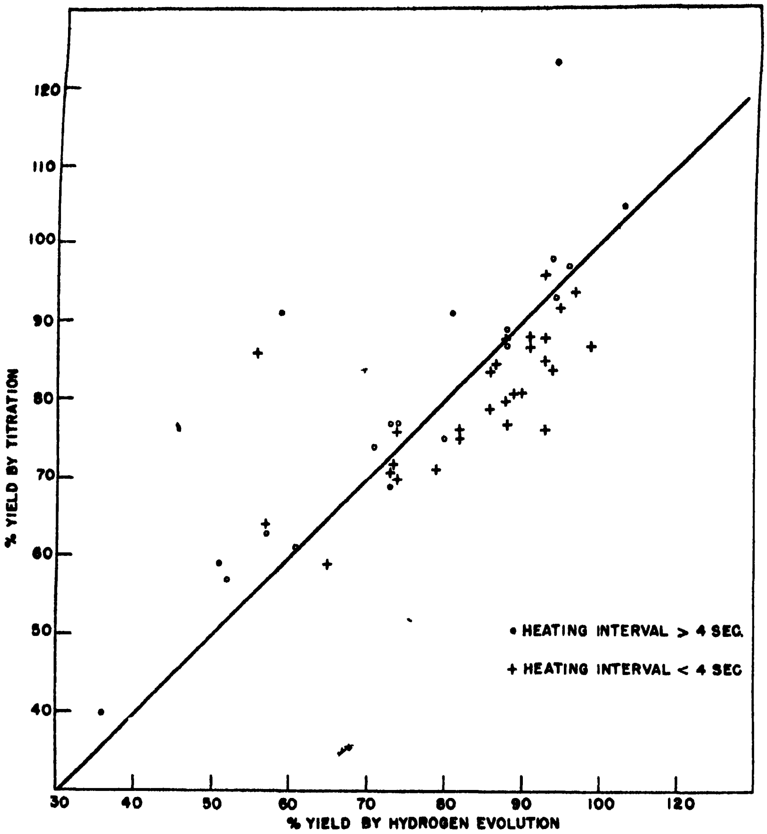


FIG. 2. Comparison of cesium yields as measured by hydrogen evolution and by titration.

Such metal would liberate hydrogen, but the corresponding cesium hydroxide would not be titrated. (Redesigning the apparatus might minimize this difficulty.) On this basis, we consider the hydrogen method to be the more reliable, provided there is no complicating liberation of excess hydrogen.

## EXPERIMENTAL RESULTS

### Silicon

The yields obtained by the hydrogen method on silicon-cesium chromate pellets of five compositions are shown in Fig. 3. The lowest abscissa plotted is 2 sec., because that is the heating interval corresponding to the maximum heating rate of which our experimental arrangement was capable. (In one instance a pellet flashed at 1 sec. and gave a 57 percent yield, as is recorded in Fig. 3. This result could not be duplicated.) It is noteworthy that the shortest heating interval becomes longer as the ratio $Si/Cs_2CrO_4$ increases; this is clear in spite of the scattering of the results, for the shortest interval at each composition was obtained with the maximum heating rate.

Figure 3 supports the following conclusions: (1) The yields fluctuate appreciably so that many measurements (more than in Fig. 3) are needed to establish all the characteristics of these pellets. (2) The theoretical yield, 100 percent, can be realized or very closely approached. (3) The yields tend to increase as the heating interval decreases. (4) The maximum yield is not highly sensitive to pellet composition in the range investigated. (If the heating rate could be increased, the yield for compositions such as $Si/Cs_2CrO_4 = 4$ would probably increase also.)

### Titanium

The striking feature of the titanium results is how frequently the calculated yields exceed 100 percent. This excess of hydrogen cannot be derived from titanium hydride, for, as has already been mentioned, the system was always thoroughly pumped out after each pellet was flashed. The most likely source of this hydrogen is the reaction of water vapor with highly reactive material, either titanium or one of its compounds, produced during the heating interval

It will be remembered that a similar evolution of excess hydrogen was observed when nickel pellet holders were used. There the reacting material was probably evaporated or sputtered metal, since the pellet holder is not likely to be involved in a chemical reaction. With titanium, however, the possibility of a lower valent compound that is a powerful reducing agent must be considered, especially since titanium is more likely to form such compounds than is silicon or zirconium.

Because of the formation of this excess hydrogen, which may have occurred no matter what the yield, there is little point in discussing the titanium results. They appear to be more erratic than those obtained with silicon, but that seems to be the principal difference. Titration ought to supplement the hydrogen method in further work with titanium; in any case, there seems to be no reason for replacing silicon by titanium in photo-tube pellets.

## Zirconium

Five experiments with $Zr/Cs_2CrO_4 = 1$ gave the following results: 82 percent at 5.0 sec., 81 per-

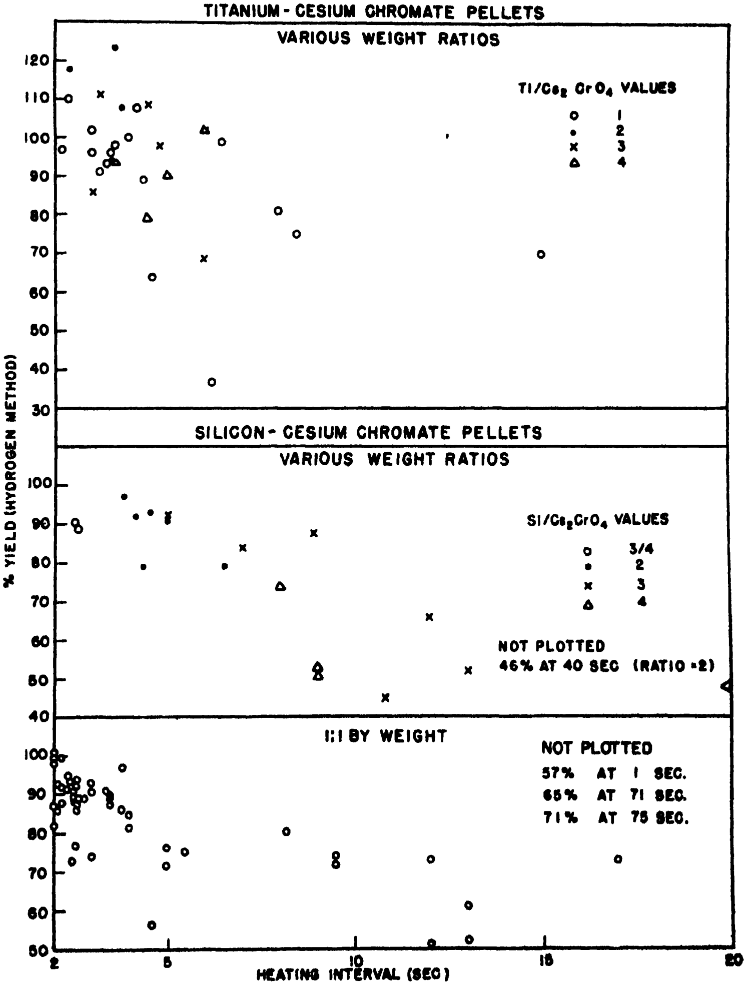


FIG. 3. Cesium yields for various compositions and various heating intervals as measured by hydrogen evolution.[1]

cent at 4.6 sec., 75 percent at 4.2 sec., 68 percent at 4.8 sec. Under our conditions, zirconium probably is a less effective reducing agent than silicon.[1]

## DISCUSSION

The present investigation is only an approach to the problem of cesium chromate photo-tube pellets, but it has shown how their yields can be measured and revealed several of their more important characteristics. The stoichiometry involved needs to be established by examining pellet residues; meanwhile, however, the equation

$$4Cs_2CrO_4+5Si=2Cr_2O_3+8Cs+5SiO_2 \quad (1)$$

may serve as a basis for a brief discussion intended primarily to point the way to future work. In reaction 1, 1 gram of silicon is equivalent to 11 grams of cesium chromate; hence, every silicon pellet investigated contained the reducing agent in large stoichiometric excess.

The heating interval may be subdivided into a relatively long initial heating period, and a short final period in which the cesium is liberated almost explosively. With the heating rate at its maximum, the heating interval lengthens as the proportion of the chromate decreases (Fig. 3). This indicates that the induction heating is being supplemented by an exothermic reaction, limited by the chromate present but evolving no cesium. For this reaction, we shall naively write

$$4Cs_2CrO_4+3Si=2Cr_2O_3+3SiO_2+4Cs_2O. \quad (2)$$

Cesium oxide would, of course, never exist free in contact with the two other oxides. But reaction 2 does emphasize that the action of hexavalent chromium on silicon supplies the heat that facilitates cesium liberation; this point of view is logical also because monovalent cesium, the other oxidizing agent present, is bound to be far the less reactive toward silicon.

For the cesium liberation we then write

$$4Cs_2O+2Si=2SiO_2+8Cs. \quad (3)$$

Actually, each silicon grain probably becomes covered, during the initial heating period, by a matrix containing the three resultants of reaction 2; cesium liberation occurs when the temperature is high enough to make possible the reaction of this matrix with silicon.

Reaction 2, if it proceeds too slowly, can give a matrix that blocks reaction 3. At least, this seems the best explanation for these observations: (1) Cesium yields tend to decrease as the heating interval increases (Fig. 3). (2) It has been observed in the factory that pellets stored near 100°C for long periods tend to give erratic yields. (3) Pellets containing dichromate, which would form more matrix since it contains more hexavalent chromium than does chromate, are more erratic than those made with chromate.

---

[1] In this connection, cf. de Boer, Broos, and Emmens, Zeits. f. anorg. allgem. Chemie, 191, 113 (1930).

# Letters to the Editor

## Remarks on Some Recently Developed Devices for Summing Fourier Series for Crystal Structure Analysis

C. L. CHRIST
*Stamford Research Laboratories, American Cyanamid Company, Stamford, Connecticut*
August 6, 1947

THERE appeared in a recent issue of this journal a paper entitled, "An electronic computer for x-ray crystal structure analysis," by R. Pepinsky.[1] The author states that the method is based upon Huggins' modification[2,3] of Bragg's earlier method[4] for photographically summing Fourier series. In the Huggins method the necessary sinusoidal distribution of bands for each Fourier term is referred to a square cell in order that the masks prepared may be used for the study of any crystal. Thus the Huggins method differs from the Bragg method in that a set of masks once prepared may be used repeatedly. With the present electronic computer the bands representing each term are oriented and spaced according to the dictates of the actual shape of the unit cell of the crystal under study.[1] This method seems, therefore, to be more closely related to the original Bragg method than to the Huggins modification.

In the electronic computer paper it is further stated[1] that the sand deposition method of McLachlan and Champaygne for summing Fourier series[5] does not permit the computation of projections having various axial ratios and angles. The fact that the "sand machine" gives projections having the same shape as that of the unit cell being studied, with regard to both the ratios of the edges of the cell and the inter-edge angle, is explicitly stated in the abstract preceeding the body of the publication cited[5] and is, of course, implicit in the detailed description of the theory of the method. In this connection it is interesting to compare the electron density projection of hexamethyl benzene given in the paper of Brockway and Robertson[6] with that obtained using the same data, with the McLachlan-Champaygne method,[5] and with the Huggins method.[3] Such a comparison shows that the use of the square lattice necessary to the Huggins method introduces distortion of the projected molecule over and above that introduced by the non-parallelness of the plane of the molecule and the plane of projection, as was pointed out by Huggins.[3] In contrast, the sand deposition method gives a projected molecule essentially the same as that obtained by Brockway and Robertson.

All three of the summation devices discussed above are based upon an equation of the type:[4]

$$\rho(x, y)' = |F_{hko}| + \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |F_{hko}| \cos[2\pi(hx+ky) - \alpha_{hko}]. \quad (1)$$

Any term $(S_{xy})_{hko}$ of this series, involving the phase angle $\alpha_{hko}$, may be represented by the sinusoidial band distribution

$$(S_{xy})_{hko} = |F_{hko}| + |F_{hko}| \cos 2\pi(hx+ky) \quad (2)$$

displaced in the positive direction along the normal to the wave crests by the linear amount $\delta_{hko}$, which may be represented by the equation

$$\delta_{hko} = \frac{\alpha_{hko}\lambda_{hko}}{360}. \quad (3)$$

Here $\lambda_{hko}$ is the distance between wave crests, and $\alpha_{hko}$ is expressed in degrees. The electronic computer, at present, is confined to projections having a center of symmetry at the origin,[1] i.e., $\alpha_{hko}$ has the value 0 or 180°. In the case of the sand deposition method $\alpha_{hko}$ may have any value whatsoever,[5] and since the true unit cell is projected, $\lambda_{hko}$ is precisely the interplanar spacing $d_{hko}$. There seems to be no reason why the Huggins method cannot also be used for non-centrosymmetric projections, since the use of Eq. (3) is compatible with the method. The desired result may be obtained, for example, in the case of non-centrosymmetric projections, by shifting the photographic film or paper on which successive exposures for each term $(S_{xy})_{hko}$ are being made[3] before each exposure by the amount $\delta_{hko}$, according to the value of $\alpha_{hko}$. It follows that for centrosymmetric cases, if this method of shift is used, only positive masks are necessary since their negative counterparts are obtained, in projection, by a shift of $\delta_{hko} = \lambda_{hko}/2$. In the case of the Huggins method it is necessary to remember that $\lambda_{hko}$ is not, in general, equal to $d_{hko}$.

---

[1] R. Pepinsky, J. App. Phys. **18**, 601 (1947).

[2] M. L. Huggins, J. Am. Chem. Soc. **63**, 66 (1941).

[3] M. L. Huggins, J. Chem. Phys. **12**, 520 (1944).

[4] W. L. Bragg, Zeits. f. Krist. **A70**, 475 (1929).

[5] D. McLachlan, Jr. and E. F. Champaygne, J. App. Phys. **17**, 1006 (1946).

[6] L. O. Brockway and J. M. Robertson, J. Chem. Soc. 1324 (1939).

# New Books

## When the Earth Quakes

By James B. Macelwane, S.J., Pp. 283, Figs. 251, 15½×23 cm. The Bruce Publishing Company, Milwaukee, 1947. Price $5.00.

This book presents in non-technical language an introduction to seismology in all its phases. This is a popular treatment of an extensive field. Neverthless the book can command the interest of the technically trained reader.

Eleven chapters are devoted to natural earthquakes. Among the subjects discussed are: kinds of earthquakes, mechanisms of earthquakes, immediate and ultimate causes of earthquakes, isostasy, earthquake prediction, field studies, instrumental studies of earthquakes, location of epicenters, sea quakes, seismic sea waves, and the principles of seismographs. Twenty modern earthquakes are described in considerable detail. Each description is accompanied by a map of the epicentral region and there are many good reproductions of photographs showing the effects of earthquakes. There is reproduced also an excellent collection of photographs of seismographs. Older types as well as the most recently developed instruments are represented.

The remaining six chapters present several less-publicized branches of seismology.

Two chapters outline the purpose, methods, and instruments of engineering seismology which is the study of buildings and structures from the point of view of resistance to damage by earthquakes. There are excellent descriptions and illustrations of accelerographs, strong motion seismographs, and vibration meters. Also explained and illustrated are the shaking machines used to generate vibrations in the ground and in buildings or other structures.

One chapter deals with seismological engineering—a term used to cover the study of industrial problems arising from man-made disturbances, such as quarry blasts or machinery vibrations, and the laboratory study of those rock properties involved in natural or artificial earthquakes.

The author describes rock bursts which constitute a serious safety problem in underground workings in some areas. He explains the rock properties which make bursts possible and shows how the seismological attack on the problem has achieved at least partial success.

In another chapter is presented an easily understood explanation of the seismic refraction and reflection methods of mapping buried structures. The machanism of petroleum accumulation is outlined and several illustrations of instruments and equipment as well as good reproductions of commercial field records.

The final chapter, entitled "Microseisms and storms at sea", is based upon recent work. The nature of the relation between large microseisms and storms at sea has been established for only a few years. The author decribes the first application of this newly acquired knowledge. During the latter part of World War II special seismograph stations were set up in the Carribean area to aid in the detection and location of storm centers at sea.

Appendix A discusses briefly the geological forces at work on the earth and explains many geological terms. There is an extensive bibliography to which references are made in the text and a glossary of terms which may be new to the average reader.

This well written book by a famous authority in seismology gives an interesting informative account of earthquake study. In addition, it does a good service for the literature by introducing aspects of modern seismology which are less familiar to workers in other fields.

The author is Dean of the Institute of Geophysical Technology of St. Louis University.

John N. Adkins
*Massachusetts Institute of Technology*

## Mechanism of Reactions at Carbon-Carbon Double Bonds

By Charles C. Price. Pp. 120, 16×23 cm. Interscience Publishers, Inc., New York., 1946. Price $2.50.

This book is based on a series of lectures presented by the author at the Polytechnic Institute of Brooklyn, and is the first volume of a series of *Lectures on Progress in Chemistry* edited by H. Mark. It is clearly pointed out in the introduction and preface that the book is not intended to be a complete detailed review of all published work on the subject, but rather a brief treatment of the pertinent information together with the author's personal interpretation of the material.

The book deals ultimately with the subject of polymerization and the various mechanisms involved. However, more than half of the book is devoted to background material on structure and reactions of unsaturated compounds in general. Modern concepts of electronic structure are presented in the first chapter together with their relation to reactivity, orientation, and reaction mechanisms. The concept of a pulsating resonant mesomerism is compared to the fixed hybrid theory of resonance, and it is apparent that the author is somewhat skeptical of certain aspects of the resonance concept as applied to organic molecules. Evidence is presented to show that an electrostatic approach may be more fundamental than the resonance hypothesis in explaining many reactions such as the orientation of substitution in benzene derivatives in which the substituent already present is one of those which do not appear to be capable of mesomeric interaction with the unsaturation of an aromatic ring or a double bond.

The second and third chapters deal with the mechanisms of ionic and free radical reactions, respectively, at double bonds. Aromatic as well as olefinic reactions are included in both chapters.

The last four chapters deal with the mechanisms involved in the various types of polymerization. The free radical mechanism of addition polymerization is discussed in Chapter 4; and the various factors involved in copolymerization are presented in Chapter 5. The author's theory that the polarity of the double bond is an important determining factor in copolymerization, is of particular interest.

Polymerization in emulsion and in suspension are treated in Chapter 6, with the function and fate of the emulsifying agent receiving particular attention. Chapter 7 concludes the book with a discussion of the polar mechanism of polymerization, and the catalysts which promote this type of reaction.

The book is well written and fully illustrated throughout with specific formulae and reactions. There are a few minor errors in the printing such as the apparent omission of certain substituent groups in formulae on pages 86 and 102. The personal manner of presentation lends interest to the book, and the author's frank appraisal of various theories and mechanisms contributes much to its value. However, the reader should keep in mind that other authorities may not be in complete accord with all of the author's interpretations. The book should be of great interest to all who are concerned with the modern concepts of organic chemistry in general, and will be of particular value to those who are concerned either with high polymer chemistry or with the general question of reaction mechanisms involving unsaturation in both aliphatic and aromatic compounds.

J. Reid Shelton
*Case Institute of Technology*

## Cosmic Radiation

Edited by W. Heisenberg. English translation by T. H. Johnson. Pp. 192, 16×24½ cm. Dover Publications, New York., 1946. Price $3.50.

We may welcome the advent of any opus edited by and contributed to by so competent a physicist as Heisenberg. Because the subject of the cosmic radiation has so few adequate, modern treatises, this one is doubly welcome. The book has come at a time to help meet the fast-growing need for good compendia in this field.

The book represents the views of a number of authors, and is a condensation of a series of lectures by these persons. Each chapter is written by a different author, the majority of whom are well-known investigators in this field. Hence the book is certainly authoritative.

The book starts with a review of the present state of cosmic-ray knowledge by Heisenberg. Next, the soft component is considered, with chapters on cascade theory by Heisenberg and on large showers by Moliere. Turning to the hard component, the bulk of the book is devoted to various phases of mesotron studies. Klemm discusses the production of mesotrons, Klemm and Heisenberg the showers with penetrating particles, Voltz the absorption of mesotrons, and Weiszacker the excitation of bursts by mesotrons. The decay is next treated by Heisenberg, the decay electrons by Bopp, and the theory of mesotrons by Weiszacker. Next follow chapters on the mesotron theory of deuterons by Flugge and the theory of explosive showers by Heisenberg. The final three chapters deal with nuclear disruptions as discussed by Bagge, the production and distribution of neutrons as treated by Flugge, and the geomagnetic effects which are summarized by Meixner. At the end a rather extensive bibliography is appended.

As would be inevitable in a work of this sort, the chapters vary somewhat in the degree of sophistication expected of the reader. Naturally, each author writes of his own specialty from his own point of view, and with varying mathematical complexity, the theoretical chapters being the most complex. The style of each author varies, but the skillful editing of the volume is evident.

The book is clearly not intended as a text, and is written primarily for the specialist in the field. It is not suitable for beginners. However, every graduate student contemplating a thesis in the subject should be thoroughly familiar with the contents. Further, since it is precisely cosmic rays themselves which provide us with our only source particles of energies higher than can be produced in man-made accelerators, the entire subject of ultra high energy physics is based on a study of cosmic-ray phenomena. Hence the book will be of value to all those interested in this rapidly expanding field. The subject is presented on the level of articles in the Reviews of Modern Physics. It will be of value to those in allied fields who wish to inform themselves about current ideas on the subject. The publication will constitute a desirable addition to any physicist's bookshelf.

S. A. Korff
*New York University*

## Modern Development of Chemotherapy

By E. Havinga, H. W. Julius, H. Veldstra, and K. C. Winkler. Pp. 175+xi, Figs. 31. Elsevier Publishing Company, Inc., Amsterdam, 1946. Price $3.50.

This volume is number four in a series of more than twenty monographs intended to record the advances made in various branches of scientific work in Holland during World War II. We cannot but be amazed at the nobility of spirit which showed itself by the perseverance in difficult activities in occupied countries during the war. This nobility is now shown again in the unique mixture of pride and generosity with which those activities are revealed to the rest of the world.

This little book summarizes the work of several Dutch laboratories up to about May 1945. In it are detailed numerous researches having to do primarily with sulfonamides, and to a very small extent with antibiotic chemotherapeutic agents. It is regrettable that this brave record can hardly be assured a permanent place in the scientific literature. Such a monograph is inevitably occupied with details too near the uncertain boundaries of knowledge to have a timeless factual value. This one is on the other hand, so limited in geographical coverage and (understandably enough) was long enough delayed in preparation that it is hardly a review of modern trends or recent advances. Especially is this true in the case of a field such as chemotherapy which is developing so rapidly all over the world.

The first and longest section deals with bacteriological investigations by Julius and Winkler on the mode of action of sulfonamides. This work is a well-conceived straightforward approach which has certainly foreshadowed many features of the theory that our more recent knowledge (particularly of the folic acids) has made more secure. Much of this approach is based upon bacterial growth curves which were probably determined more reliably than in almost any similar study. Modifying the environments of the bacteria gave the opportunity to study the intimate effects of various inhibitors and activators. Wood's *para*-aminobenzoic acid displacement theory is re-examined with an admirable detachment. The authors insist that an antagonistic displacement of *para*-aminobenzoic acid by sulfonamide has never been demonstrated, only the reverse antagonism being known. The net result is that sulfonamides are viewed as interfering with a mechanism (involving in some way *para*-aminobenzoic acid) which leads to the synthesis of a peptone-like substance.

Here and elsewhere in the book, the authors discuss certain of the findings of workers elsewhere that were available to them through abstracts or otherwise. This is true even though in the preface it is stated that "foreign work will be referred to only in so far as it has a direct bearing on Dutch research work."

A section describing physico-chemical investigations on the sulfonamides by Havinga and Veldstra does not materially increase our knowledge of the mode of action of these substances. Certain studies were made of the titration curves, light absorption spectra, polarographic behavior, and the properties of surface films. The impression is given that these approaches were chiefly exploratory, based upon plausible but casual hypotheses. An interesting part of this section is the attempt to estimate the collision frequency of *para*-aminobenzoic acid or sulfonamide molecules with receptor spots upon a bacterial cell. If it is assumed that the former substance is required only at certain moments of the division cycle for normal growth to occur, then it may be easier to understand why the antagonists must be present in much higher concentration if they are to prevent growth.

The results of the testing of approximately seventy-five synthetic compounds related to sulfanilamide are given. There is a brief summary of clinical and animal testing of a few of these substances. Finally, the mold antibiotic "expansine" is described, a substance that appears to be identical with clavacin (and patulin) discovered elsewhere at about the same time.

This volume makes available in very satisfactory English translation a summary of brave and independent investigations.

Rollin D. Hotchkiss
*Rockefeller Institute*

## High Vacuum Technique

By J. Yarwood. Second Edition. 15½×24 cm. John Wiley and Sons, Inc., New York, 1945. Price $2.75.

From the authors intention, which, as stated in the preface, is "to describe important industrial processes hitherto discussed only in scientific periodicals," it is difficult to discern to whom he addresses this addition to the "know-how" literature.

The author's intention "to introduce the latest developments . . . with special emphasis on the most recent work," is unfulfilled—the author does not, for example, discuss the palladium-window ion gauge for leak hunting, the silicone-diffusion pump oils, or Vycore.

There are instances where the author contributes but little to the subject he treats—for example, on page 82, after stating that silver oxidation can be followed by surface color changes the exposition concludes with: "The degree of oxidation is found only by experience based on previous experimental work . . . ."

The judgment of your reviewer does not always concur with that of the author on the value of the apparatus or of the procedures described and evaluated. It would be helpful to know which procedures the author knows first hand and intimately, and which he is merely re-writing, or reporting.

The author occasionally lets himself speak in the vernacular as: "The gun must be vacuum-stoved before mounting in the tube." In a book such as this the use of the vernacular seldom enlightens those who understand it and is unintelligible to those one would teach.

The chapter on the properties of materials does contribute to the author's stated intention, "to bring together in one volume the diverse facts regarding the relevant properties and uses of materials encountered in all types of vacuum work." However, his treatment of some materials, like copper, is certainly not up-to-date from the high vacuum point of view.

It is to be expected that engineers and experimental physicists will get substantial help and provocative suggestions from Yarwood's work. The references to the literature will be of value to some; several procedures are described which will be novel to others; and the subject matter is more modern than the "classics", to which the author refers.

John D. Strong
*The Johns Hopkins University*

## Electrical Transmission in Steady State

By Paul J. Selgin. Pp. 427, Figs. 100, 14½×22 cm. McGraw-Hill Book Company, Inc., New York, 1946. Price $5.00.

This book, like many others recently published, is the outgrowth of notes developed in connection with courses taught for war-training purposes. For this reason the manner of presentation of the subject matter, as well as its selection, is not entirely what one would desire for ordinary undergraduate instruction in an electrical communications curriculum on a peacetime basis. The book is essentially one on the subject of transmission lines for use at high and ultra high frequencies, the first 332 pages out of a total of 416 being devoted to this topic. Following this material there is a chapter on inductively coupled circuits, a ten-page chapter on capacitive coupling, and two chapters, totalling 34 pages, on high frequency vacuum-tube amplifiers.

The first two chapters comprise a condensed review of some basic lumped-constant circuit principles, including such topics as matching and optimizing conditions for power transfer at junctions, transmission and insertion loss, reciprocity, Thévenin's and Norton's theorems, image parameters, etc. The steady-state behavior of transmission lines is then dealt with through considering the limit approached by an increasing number of smaller and smaller lumped networks in cascade, and a detailed (rather involved) discussion is given of the frequency behavior of the attenuation and phase functions. At this point and throughout the remainder of the book, one has a very definite conviction that a student having no previous acquaintance with this subject would find the going very rugged indeed, unless considerable additional guidance were available from an experienced instructor.

The input impedance of the line is dealt with through a discussion of the mapping properties of the hyperbolic tangent function, followed by a consideration of the impedance transformations of analogous lumped networks based upon the circle properties of simple inversion. The bi-polar line chart is then introduced and applied to various conventional impedance-transformation problems, including the quarter-wave line as an impedance transformer. This discussion is generalized to a consideration of cascades of quarter-wave lines used to secure a more frequency insensitive impedance transformation, and an "optimum" design of such a transformer is given which, after several pages of involved algebra, finishes with the statement of an end result, the proof of which "although not particularly difficult, is beyond the scope of the present discussion" (footnote, page 203 . Single and double stub tuners are discussed with the help of the Smith Chart (not so called), and detailed treatment is given of the frequency sensitivity of such tuners, taking into account the effect of incidental dissipation and expressing these effects in terms of the "$Q$" of the system.

Chapter XI then turns to a consideration of a review of static field theory, a discussion that in the next chapter is extended to dynamic fields and Maxwell's equations. This material admittedly is too condensed to be helpful to the beginner, while for the reader with sufficient background to follow the argument, it fails to provide a picture that is compact and effective in presenting the essential relationships clearly and completely. While a reasonably good job is done here for readers who at this stage do not expect to acquire a proficient knowledge of these tools, there is left much room for improvement.

After some discussion concerning the relationship of lumped circuit theory to field theory, the author considers the propagation of plane waves and derives the uniform line equations from the field equations, extending this discussion to the consideration of the exponential tapered line and the exponential stub line.

The subject of inductively coupled, lumped tuned circuits given in the following chapter (XIV), and introduced through a consideration of self- and mutual inductances and the simple transformer, is dealt with in a manner as involved as any that this reviewer has seen to date (the average student must surely despair at this point). The succeeding short chapter on capacitive coupling is given primarily as a means for dealing with the inter-electrode capacitances of vacuum tubes, of which some basic principles are discussed in the concluding brief pages of this book.

A feature on which the author is to be complimented is the inclusion throughout the book of illustrative numerical examples fully worked out.

In view of some of the above rather harshly critical comments it is only fair to state in conclusion that for the purpose for which this material was originally intended, the accomplishment of the author ranks on the whole with the better text material of a similar nature. Now that war-training programs are over, however, it is to be hoped that more attention in textbook writing will be given to the needs of students following a more orderly and progressively stimulating program of education in basic circuit theory and closely related subjects.

E. A. GUILLEMIN
*Massachusetts Institute of Technology*

# New Booklets

Fischer and Porter Company, Hatboro, Pennsylvania, has issued a 32-page booklet entitled *A New Era in Liquid and Gas Flow Measurement*, which describes the advantages of the flowrator meter for flow rate measurement. Address request to Department 6F-F of the company.

Burrell Technical Supply Company, 1936 Fifth Avenue, Pittsburgh 19, Pennsylvania, has published a bulletin on its *Junior Spectrophotometer--The Successor to Colorimeters*. 8 pages, free on request.

The Daven Company, 191 Central Avenue, Newark 4, New Jersey, announces the publication of a new bulletin of technical data, presenting information widely used in the design of sound equipment. 4 pages. Free on request. Write on company letterhead.

Tin Research Institute, Fraser Road, Greenford, Middlesex, England, issues at regular intervals (because of present conditions) a review entitled *Tin and Its Uses*. The issue of July 1947 contains 16 pages and has numerous photographs in color. Interested readers are invited to address a request to the Publications Department of the Institute, stating that they wish to receive, free of charge, any publications of the Tin Research Institute dealing with given subjects or products (naming them), and mentioning their own business connections and titles. The July 1947 issue may be obtained directly from the Battelle Memorial Institute, 505 King Avenue, Columbus, Ohio.

The Gaertner Scientific Corporation, 1201 Wrightwood Avenue, Chicago 14, Illinois, announces two new bulletins: 140-74, entitled *Dilatation Interferometer*, and an unnumbered booklet entitled *Survey of Precision Instruments*, which is intended to serve as a guide to the company's products. Both bulletins are eight pages in length and are free on request.

*Interchemical Review* for Summer 1947, published by the Interchemical Corporation, 432 West 45th Street, New York 19, New York, features "Acids, bases, salts—pH," by C. Ross Bloomquist, and "Printing inks for postage stamps," by Milton Zucker. 30 pages.

Eberbach and Son Company, Ann Arbor, Michigan, has published the fall issue of its house magazine, the *Announcer of Scientific Equipment*. The feature article is entitled "Functional design of laboratory glassware". Free on request.

The National Bureau of Standards has announced the publication of the following bulletins in the physics field. All are available at 10 cents per copy from the Superintendent of Documents, Government Printing Office, Washington 25, D. C.

RP 1807—*Solution of Problem of Producing Uniform Abrasion and Its Application to the Testing of Textiles*" by Herbert F. Schiefer.
RP 1808—*Chromaticities of Lovibond Glasses* by Geraldine W. Haupt and Florence L. Douglas.
RP 1813—*Resistance-Temperature Relation and Thermoelectric Properties of Uranium* by Andrew I. Dahl and Milton S. Van Dusen.
RP 1814—*Infared Absorption Spectrum of Carbon Disulfide* by Earle K. Plyler and Curtis J. Humphreys.

North American Philips Company, Inc., 100 East 42nd Street, New York 17, New York, has issued a 4-page folder entitled *X-Ray Diffraction Camera for Microtechniques.* The company states that it should be of value to those actively engaged in fiber research. Free on request.

The *RCA Review* Department of RCA Laboratories Division, Princeton, New Jersey, has published the first volume of its new *Engineering Book Series*, entitled *Patent Notes for Engineers.* Copies may be obtained from *RCA Review* at $2.50 per copy for the permanent clothbound edition (plus 20 cents per copy postage if mailing address is outside the United States). The book was written by the director of the company's patent department.

# Here and There

### New Appointments

Bruce S. Old has been appointed consultant to the Atomic Energy Commission, Division of Research, in the capacity of chief metallurgist. He will continue his present work with Arthur D. Little, Inc., on a part-time basis. Dr. Old will assist in organizing the work of the Commission in the metallurgical field.

Irving A. Denison has been named Chief of the Underground Corrosion Section of the National Bureau of Standards.

Bowen C. Dees (until recently an Assistant Professor of Physics at Rensselaer Polytechnic Institute) has accepted a position as physicist with the Economic and Scientific Section, under the Supreme Commander for the Allied Powers in Tokyo. He will survey and advise concerning the physical research being conducted in the university and commercial laboratories in Japan.

Ernest R. Hanson has joined the staff of Foster D. Snell, Inc., to direct research and development in the field of plastics and rubber.

### Necrology

Dr. Selig Hecht, Professor of Biophysics at Columbia University, died in New York on September 18. A distinguished authority in the field of vision, Dr. Hecht had been an Associate Editor of the Journal of Applied Physics since 1943.

### Awards

Warren A. Marrison, a member of the technical staff of Bell Telephone Laboratories, has been awarded the British Horological Institute's gold medal for 1947 in recognition of pioneer researches in the development of the quartz crystal clock, which has become the most accurate timekeeper in the world.

A. A. Knowlton, veteran physics professor at Reed College, Portland, Oregon, was recently voted a special award of $1,000 for distinguished teaching by directors of Research Corporation. The award is unique in that Dr. Knowlton, now in his thirty-third year at Reed, may spend it in any way he wishes. It was based on the extraordinary achievement of Dr. Knowlton's graduates. Specifically, the award stems from a recent study in the American Journal of Physics which showed that of all physicists granted Ph.D. degrees by American universities between 1936 and 1945, a disproportionate number had done their undergraduate work at one of two small colleges—Reed, and North Central College, Naperville, Illinois. (A similar award has been voted to Dr. Clifford N. Wall, formerly of North Central.)

### Lincoln Arc Welding Foundation Award and Scholarship Program

The James F. Lincoln Arc Welding Foundation, Cleveland 1, Ohio, has issued a brochure outlining the continuing features of its ten-year award and scholarship program designed to help develop the arc welding industry. Under the award plan, engineering students of various colleges will submit papers on arc welded design. Under the scholarship plan, scholarships will be allocated to the schools in which the three highest award winners under the award program are registered. Awards will total $5,000 and scholarships, $1,750. Rules for the 1947-1948 competition may be obtained by addressing the Secretary of the Foundation at Cleveland 1, Ohio.

### Research Corporation Reports on Grants

Research Corporation, which in late 1945 established the Frederick Gardner Cottrell Fund for grants in aid of postwar science, has announced that in the following year more than 100 grants were approved, totaling about $430,000. The grantee institutions are located in about thirty states, Alaska, and the District of Columbia. The awards were made largely to aid younger men just returned to academic research from wartime occupations.

### Conference on the Administration of Research

A conference on the administration of research was held at The Pennsylvania State College October 6 and 7, 1947. It was sponsored by the School of Engineering of the College. Two hundred people registered, representing laboratories, both large and small, managed by industry, educational institutions, and governmental agencies. Proceedings of the conference will be published as soon as possible, at a price of $3. Persons who wish to have a copy and who did not have an opportunity to ask for it at the conference may order from Professor Kenneth L. Holderman, The Pennsylvania State College, State College, Pennsylvania.

### Pittsburgh Physical Society Officers

The Physical Society of Pittsburgh has elected the following officers for the year 1947–1948: *President*, Dr. Frank Morgan, Gulf Research and Development Corporation; *Vice President*, Dr. E. Stickley, Pittsburgh Plate Glass Company; *Secretary-Treasurer*, Dr. F. A. Schwertz, Mellon Institute.

### Bureau of Standards Establishes Atomic Physics Division

The National Bureau of Standards has formed a Division of Atomic Physics in which the main Bureau activities relating to atomic and molecular physics have been grouped together. The functions of the new division are the promotion of fundamental fact finding research and the precise determination of important fundamental standards in the field of atomic physics. Six sections make up the new division: spectroscopy, electronics, mass spectrometry, radioactivity, x-rays, and atomic physics. Dr. E. U. Condon, Director of the Bureau of Standards, has assumed charge of the new division, and Dr. Robert D. Huntoon has been named assistant chief.

### Conference on the Use of Radioactive Isotopes in Agricultural Research

The Alabama Polytechnic Institute, in cooperation with the Oak Ridge Institute of Nuclear Studies, sponsored a three-day conference on the Use of Radioactive Isotopes in Agricultural Research at Auburn, Alabama, December 18–20, 1947. The attendance, through special invitations, consisted of representative scientists largely from the southeastern States.

### Lalor Foundation Fellowship Program

The Lalor Foundation fellowship program at the Marine Biological Laboratory, inaugurated in 1947, will continue in 1948. Postdoctoral summer fellowships in the fields of biophysics, biochemistry, and physiological chemistry will be available. The fellowships are designed primarily to aid promising young investigators who can make maximum use of the facilities and opportunities provided at the Marine Biological Laboratory. In addition to laboratory facilities, the grants are intended to cover approximately the living expenses of the investigators at Woods Hole and necessary traveling expenses to and from Woods Hole. The committee administering the fellowships consists of Drs. Eric Ball, Kenneth Cole, Charles Packard, A. K. Parpart, and C. W. Metz (Chairman). Inquiries and applications should be addressed to the Director of the Marine Biological Laboratory, Woods Hole, Massachusetts. The applications should be received by December 31, 1947, but will be considered after this date.

### Calendar of Meetings

**December, 1947**

29–31 American Physical Society (and its Division of High-Polymer Physics), American Association of Physics Teachers, *and* the American Association for the Advancement of Science, Chicago, Illinois

**January, 1948**

2– 3 American Physical Society Los Angeles, California
21–23 American Society of Civil Engineers, New York, New York
29–31 American Physical Society, New York, New York

**March**

1– 5 American Society for Testing Materials (Spring Meeting), Washington, D. C.
22–25 Institute of Radio Engineers, New York, New York

**April**

19–23 American Chemical Society, Chicago, Illinois
22–24 Acoustical Society of America, Washington, D. C.

# Author Index to Volume 18

References with (L) are to Letters to the Editor.

# Subject Index to Volume 18

**References with (L) are to Letters to the Editor.**